



***Facultad
de
Ciencias***

**Optimización de la selección de eventos
ttbar con HF-jets adicionales en el detector
CMS**

**(Optimization of ttbar event selection with
additional HF-jets in the CMS detector)**

**Trabajo de Fin de Grado
para acceder al**

GRADO EN FÍSICA

Autor: Pablo Herreros Fuentevilla

Director: Javier Andrés Brochero Cifuentes

09 - 2020

Resumen

En este proyecto, se han estudiado eventos $t\bar{t}$ (con jets adicionales) en el canal de decaimiento semileptónico, con el objetivo de optimizar su selección. Esta optimización se basa en la creación de dos nuevas variables a partir del método de etiquetado de b-jets que se utiliza en el detector CMS del acelerador LHC. Para analizar los resultados, se han estudiado y medido las secciones eficaces de los procesos $t\bar{t}$, $t\bar{t}b\bar{b}$ y $t\bar{t}H$. Para ello, se han utilizado datos de colisiones protón-protón a una energía de centro de masas de $\sqrt{s} = 13$ TeV tomados durante el año 2017. El análisis de los tres procesos muestra una mejoría en la selección de eventos manteniendo las secciones eficaces compatibles con los valores teóricos predichos por el modelo estándar.

Palabras clave: quark top, jets, etiquetado b-jets, CMS, sección eficaz, bosón de Higgs

Abstract

In this project is studied $t\bar{t}$ events (with additional jets) in the semi-leptonic decay mode, aiming to optimize their event selection. This optimization is based on the creation of two new variables from the b-jets tagging method used in the CMS experiment at the LHC particle accelerator. In order to analyze the results, cross sections of $t\bar{t}$, $t\bar{t}b\bar{b}$ and $t\bar{t}H$ processes are studied and measured. For that purpose, it is used data collected in proton-proton collision at centre of mass energy of $\sqrt{s} = 13$ TeV taken during 2017. The analysis of these three processes shows an improvement in the event selection maintaining cross sections compatible with the theoretical values predicted by the Standard Model.

Keywords: top quark, jets, b-tagging, CMS, cross section, Higgs boson

Agradecimientos

Quiero dar las gracias a mi tutor en este trabajo, Javier Brochero, Post-Doc Juan de la Cierva (universidad de Cantabria), por sus consejos y guía en la realización de este trabajo. Agradezco mucho su paciencia y que siempre sacara tiempo para ayudarme con mis dudas.

También quiero agradecer a mi familia y amigos el apoyo incondicional durante todos estos años.

Índice

1. Introducción	5
2. Marco Teórico	6
2.1. El modelo estándar	6
2.2. Física del quark top	8
2.2.1. Producción del quark top	8
2.2.1.1. Producción eventos $t\bar{t}$	8
2.2.1.2. Producción quarks top aislados	10
2.2.2. Estados finales para eventos $t\bar{t}$	10
2.2.2.1. $t\bar{t}$ en canal semileptónico	12
2.2.2.2. Fondos en el canal semileptónico	12
3. Experimento LHC	16
3.1. Introducción	16
3.2. CMS	17
3.2.1. Sistema de coordenadas	18
3.2.2. Tracker	19
3.2.3. Calorímetros	19
3.2.4. Detector de muones	19
3.3. b-tagging	20
4. Análisis	22
4.1. Objetos	22
4.2. Categorización de eventos $t\bar{t}$ + jets	23
4.3. Gráficos de control	24
4.3.1. Regiones de ≥ 4 jets y ≥ 6 jets	24
4.3.2. Región para eventos $t\bar{t}$	25
4.3.3. Regiones para eventos $t\bar{t}$ +jets	31
4.4. Construcción de variables para la identificación de eventos con b-jets	32
4.5. Distribuciones de ≥ 4 jets y ≥ 6 jets para las nuevas variables	33
4.6. Eficiencia y significancia	34
4.7. Cortes aplicados a las distribuciones de ≥ 4 jets y ≥ 6 jets	36
4.8. Pureza y sección eficaz de $t\bar{t}$	37
4.9. Pureza y sección eficaz de los procesos $t\bar{t}b\bar{b}$ y $t\bar{t}H$	38
5. Conclusiones	39
6. Bibliografía	41

1. Introducción

Entre los años 2016-2018 en el detector *Compact Muon Solenoid* (CMS) del *Large Hadron Collider* (LHC) se han recogido una gran cantidad de datos de colisiones protón-protón a una energía de centro de masas de $\sqrt{s} = 13$ TeV. En este trabajo se han utilizado los datos recogidos en el 2017, en este mismo experimento, correspondientes a una luminosidad integrada de $\mathcal{L} = 41.5 \text{ pb}^{-1}$.

En el año 1995 fue descubierto, en el colisionador Tevatron, el quark top cuyo estudio es fundamental para la validación del modelo estándar ya que permite realizar mediciones muy precisas. Además, a causa de las propiedades características de este quark, entre las que destaca que decae antes de hadronizar, el estudio de esta partícula puede llevar al descubrimiento de nueva física. Algo similar sucede con el bosón de Higgs, que debido a su reciente descubrimiento (2012) se trata de una partícula cuyo estudio resulta muy interesante para probar y ampliar el conocimiento actual acerca del modelo estándar (SM). El proceso en el que se centra el trabajo es la producción de eventos de pares top anti-top en el modo de desintegración semileptónico. Se utiliza este modo de desintegración debido a su alta estadística. Un caso particular, es la producción de pares top anti-top junto a dos b-jets adicionales ($t\bar{t}b\bar{b}$) el cuál es el principal fondo para las distintas mediciones del bosón de Higgs cuando este decae a dos quarks b ($t\bar{t}H$). Separar el proceso $t\bar{t}b\bar{b}$ del proceso $t\bar{t}H$ es todavía un desafío.

En este trabajo, se propone una optimización del método de etiquetado de b-jets que se utiliza en el CMS (DeepCSV), basado en la creación de dos nuevas variables de tal forma que se pueda mejorar la selección de este tipo de eventos. Adicionalmente, también se presentan las medidas de las secciones eficaces para los procesos $t\bar{t}$, $t\bar{t}b\bar{b}$, $t\bar{t}H$.

Este trabajo se divide en tres partes bien diferenciadas. En primer lugar, se presenta un marco teórico que va desde el modelo estándar pasando por las propiedades del quark top hasta las características de los procesos que se van a estudiar y sus posibles fondos. Después se explica el experimento LHC, concretamente el detector CMS y sus subdetectores, así como el proceso de etiquetado de b-jets (*b-tagging*). Por último, se realiza el análisis de los datos recogidos así como sus resultados.

2. Marco Teórico

2.1. El modelo estándar

En física de partículas, el modelo estándar describe las partículas fundamentales que forman la materia así como sus interacciones. Por un lado, las partículas fundamentales se dividen en dos grupos, en función de su *spin*. Si tienen un *spin* semientero se les denomina fermiones mientras que si el *spin* es entero son bosones. Dentro de los fermiones se distinguen dos conjuntos distintos de partículas, los quarks (q) y los leptones (ℓ). Ambos grupos se dividen en tres generaciones, formadas por pares de partículas, donde la generación I corresponde con las partículas con menor masa y más estables hasta llegar a la generación III donde se encuentran las partículas mas inestables y pesadas. De esta manera, los quarks tienen 6 ‘sabores’ distintos: *up* (u), *down* (d), *strange* (s), *charm* (c), *top* (t) y *bottom* (b). Según su carga eléctrica, se dice que los quarks u , c , t son tipo- u ya que tienen los tres una carga eléctrica de $\frac{2}{3}e$ mientras que los quarks d , s , b son tipo- d ya que su carga eléctrica es de $-\frac{1}{3}e$. Por su parte, los leptones tienen otros seis sabores distintos: electrón (e), muón (μ), tau (τ), neutrino electrónico (ν_e), neutrino muónico (ν_μ) y neutrino tau (ν_τ). Los electrones, muones y tau tienen carga $-1e$ mientras que sus respectivos neutrinos carecen de carga eléctrica y su masa es muy pequeña. Todos estos fermiones tienen su correspondiente antipartícula, la cuál posee la misma masa pero con números cuánticos opuestos.

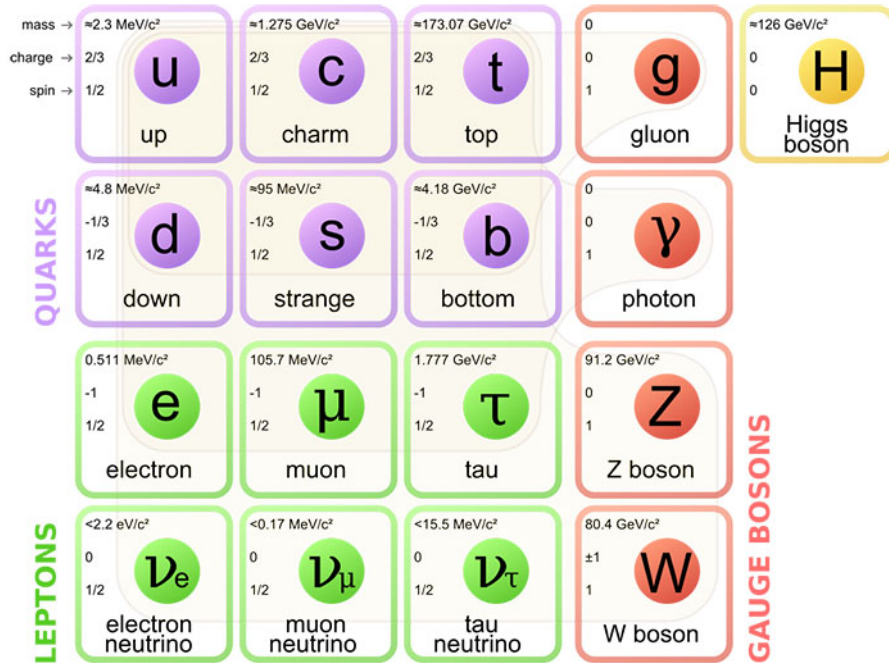


Figura 1: Modelo estándar de partículas.[2]

Si la partícula tiene *spin* entero, se la denomina bosón. Los bosones (*spin*=1) son los responsables de las interacciones entre partículas. En la naturaleza hay 4 fuerzas fundamentales: la fuerza electromagnética (EM), la fuerza fuerte, la fuerza débil y la fuerza gravitatoria. Esta última fuerza, no está incluida en el SM, ya que como solamente domina a partir de masas muy grandes comparadas con la de estas partículas, no es necesario considerarla en el estudio de estas interacciones. La inclusión de esta fuerza en el modelo estándar es un problema abierto en la física actual. En cambio, las otras tres interacciones si que están explicadas a través de este modelo y sus bosones.

La fuerza electromagnética tiene como mediadores los fotones, que son bosones sin carga eléctrica ni masa. Como no tienen masa, esto hace que su rango de interacción sea infinito.

Después se encuentra **la fuerza débil**, la cuál es la responsable del decaimiento nuclear beta. Sus mediadores son los bosones W^\pm y el bosón Z. La diferencia entre estos bosones es que los W^\pm tienen carga eléctrica mientras que el Z no, además de que el Z no cambia el sabor de las partículas. Estos bosones, al contrario que los fotones, tienen masa y por lo tanto su alcance es finito del orden de $\sim 10^{-3}\text{fm}$ [1].

Por último, se encuentra **la fuerza fuerte**, que es la que se encarga de mantener el núcleo atómico estable. Tiene un rango limitado (1 fm) [1] y las partículas mediadoras son los gluones, que no tienen carga eléctrica pero si carga de color.

Interacción	bosón	spin	masa[GeV]	fuerza relativa	carga eléctrica	carga de color
Fuerte	gluon (g)	1	0	1	0	si
Débil	W^\pm	1	80.4	10^{-13}	± 1	no
Débil	Z	1	91.2	10^{-13}	0	no
Electromagnética	fotón (γ)	1	0	10^{-3}	0	no

Tabla 1: Fuerzas fundamentales de la naturaleza [3]

Además de estos bosones de *spin* 1, existe un bosón escalar (*spin* 0) que explica que la masa de los bosones W^\pm y Z no sea nula. Se trata del recientemente descubierto (2012) bosón de Higgs, que explica mediante el mecanismo de Higgs la masa de estas partículas.

Los quarks se pueden agrupar, vía fuerza fuerte, formando bariones (3 quarks) o mesones (quark-antiquark). A causa del confinamiento de color, solo pueden ser observados en estados sin color. Esto implica que en la naturaleza no se pueden encontrar aislados. Si se intentan separar se produce lo que se conoce como hadronización, es decir, la densidad de materia aumenta a medida que se separan dando lugar a la formación de otra serie de partículas, sin carga de color, como pueden ser los hadrones o piones. Esto da lugar a lo que se conocen como jets, que se define como el conjunto de partículas creadas a partir de la hadronización con un mismo origen [3].

2.2. Física del quark *top*

El quark *top*, como se ha mencionado anteriormente, es un quark tipo u y pertenece a la generación III de quarks, junto al quark *bottom*. Esto ya deja entrever muchas de sus propiedades, como la carga eléctrica o la masa. Se trata del quark con más masa y por lo tanto es el menos estable. Pero esto no es lo único, ya que además es la partícula más pesada del modelo estándar. La masa del quark top está definida experimentalmente por la posición del pico de la distribución de la masa invariante de sus productos de decaimiento (bosón W y un b jet). El valor medio (mundial) obtenido en la determinación de la masa del quark t es $m_t = 173.34 \pm 0.27(\text{stat.}) \pm 0.71(\text{syst.}) \text{ GeV}$ [4]. Es por esta característica que esta partícula tiene un gran interés. Debido a su masa, el quark *top* no pudo ser observado directamente en el LEP y fue descubierto en 1995, en el Tevatron, mediante colisiones proton - antiprotón ($p\bar{p}$).

El quark t es el único que decae mediante interacciones débiles de primer orden como un bosón W y un quark, ya que su masa es mucho mayor que la del bosón W^\pm . En ese vértice, el quark más probable es el b porque para el s o el d los factores de la matriz CKM son muy pequeños. Además, como su tiempo de vida es muy corto debido a su decaimiento, es menor al tiempo necesario para que la hadronización pueda suceder y por lo tanto no se pueden encontrar hadrones con quarks t o jets que tengan su origen en estos.

La física del quark *top* permite poner a prueba nuestro conocimiento acerca de las interacciones fuertes así como a través del conocimiento de sus propiedades se puede obtener información muy importante de la ruptura de simetría en la teoría electro-débil.

2.2.1. Producción del quark top

En el modelo estándar, los dos principales mecanismos de producción de quarks *top* en los colisionadores de hadrones son la producción de pares *top* anti-*top* mediante la interacción fuerte (el más dominante) y la producción de quarks *top* aislados a través de la interacción electro-débil.

2.2.1.1. Producción eventos $t\bar{t}$

La producción de pares top anti-top puede ser bien mediante colisiones quark anti-quark como a partir de la fusión de gluones. En el LHC, donde se hacen colisionar *bunches* de protones, la producción de pares top anti-top mediante las aniquilaciones quark anti-quark representan un 17 %, mientras que el proceso dominante es la fusión de gluones con un 83 %.

Para estudiar la probabilidad de que sucedan un determinado proceso, en física de partículas se introduce el concepto de *cross section* (sección eficaz). De una forma muy general, se puede describir como la probabilidad de que dos partículas choquen y reaccionen de una determinada manera. En el caso de querer estudiar la producción de un par top anti-top, la sección eficaz sería el número de eventos en los se crea un par top anti-top entre el número de colisiones que se han producido con los protones utilizados. La unidad de medida de esta variable son los *barns* (b) donde $1\text{b}=10^{-24} \text{ cm}$. Esta variable es de mucha utilidad ya que es independiente de la intensidad de los haces de partículas por lo que se pueden comparar los

resultados obtenidos en distintos detectores [5].

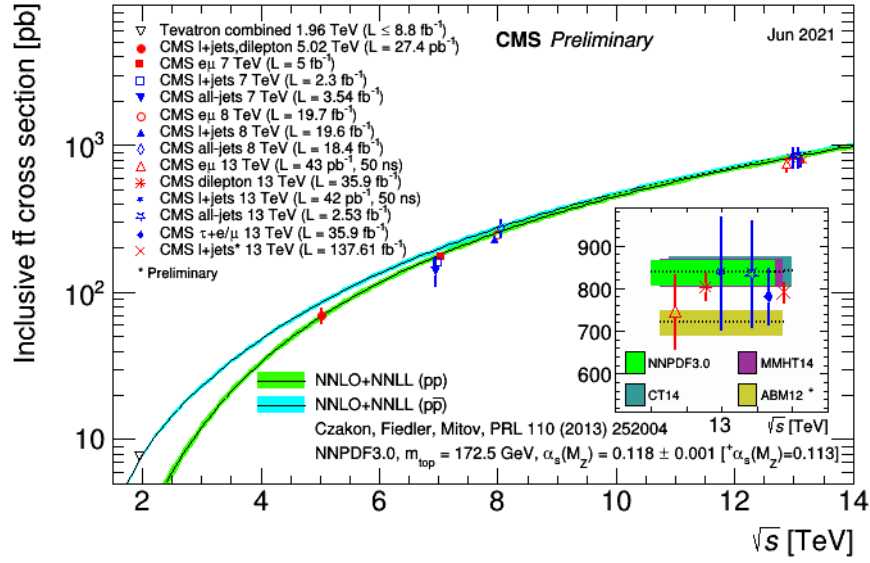


Figura 2: Sección eficaz de la producción del quark top.[6]

En la figura 2, se puede observar un gráfico correspondiente a la medición de la *cross section* de los eventos $t\bar{t}$. Aquí se aprecia como varía en función de la energía del centro de masas que se está utilizando. Las líneas continuas corresponden a la predicción teórica, que son dos porque se han utilizado colisiones protón-protón y colisiones protón-antiprotón. La línea verde corresponde con las colisiones pp y la azul con los $p\bar{p}$. Principalmente los puntos experimentales se han medido a unas energías de centro de masas de 7, 8 y 13 TeV, que son las energías a las que ha funcionado el LHC a lo largo de los años. Hay otras dos medidas que corresponden a menores energías, una de ellas realizada en el Tevatron. En la parte derecha, se puede ver la zona de los puntos experimentales ampliada, observando así el valor que toman y se comprueba que son compatibles con la predicción teórica.

Energía[TeV]	sección eficaz[pb]	Scale[pb]	PDF[pb]
7	177.31	+4.56 -5.99	+9.02 -9.02
8	252.89	+6.39 -8.64	+11.67 -11.67
13	831.76	+19.77 -39.20	+35.06 -35.06
14	984.50	+23.21 -34.69	+41.31 -41.31

Tabla 2: Sección eficaz $t\bar{t}$ para una masa de 172.5 GeV [7]

En la Tabla 2, se han presentado los distintos valores de la *cross section* en función de la energía de centro de masas, junto a sus incertidumbres teóricas, considerando que la masa del quark *top* es de 172.5 GeV.

2.2.1.2. Producción quarks top aislados

La producción de quarks top aislados puede ser mediante la interacción electro-débil involucrando un vértice Wtb , donde se distinguen tres modos distintos [8]:

1. *t-channel*: Este proceso es el más dominante. Consiste en que un bosón W virtual interactúa con un quark b que se encuentra en el interior del protón. A este modo también se lo conoce como fusión de gluones donde el quark b se produce a partir de un gluón que se divide en un par $b\bar{b}$.
2. *s-channel*: Este modo de producción es del tipo Drell-Yan y es el menos dominante de los tres. Consiste en que la fusión de dos quarks producen un bosón W .
3. *Associated production (tW)*: El quark top es producido junto a un bosón W real. El quark b inicial se encuentra en el interior del protón.

A continuación se muestran los valores de la *cross section* para los procesos *Single Top* (un solo quark top), únicamente para el *t-channel*.

Energía [TeV]	sección eficaz [pb]	Scale [pb]	PDF [pb]	Total unc. [pb]
7	63.89	+1.92 -1.25	+2.19 -2.19	+2.91 -2.52
8	84.69	+2.56 -1.68	+2.76 -2.76	+3.76 -3.23
13	216.99	+6.62 -4.64	+6.16 -6.16	+9.04 -7.71
14	248.09	+7.58 -5.40	+6.98 -6.98	+10.30 -8.82

Tabla 3: Sección eficaz single top en el *t-channel*. [9]

2.2.2. Estados finales para eventos $t\bar{t}$

Como se ha mencionado anteriormente, casi todas las veces (99.8 %) el quark top decae en un bosón W y un quark b . Esto implica por lo tanto, que en los procesos $t\bar{t}$, estos van a decaer en dos bosones W y dos quarks b , los cuáles serán detectados como jets. Los dos bosones W pueden decaer en distintos tipos de partículas (leptones o hadrones) de tal forma que se pueden distinguir distintos modos de decaimiento.

1. Por un lado está el decaimiento **dileptónico**, que consiste en que ambos bosones W decaigan en electrones o muones.
2. Si un bosón W decae en un electrón o un muón mientras que el otro decae en quarks, se trata de un modo **semileptónico**.
3. En el caso de que ambos bosones W decaigan en quarks se tratan de un decaimiento **completamente hadrónico**.
4. Por último, se encuentra el caso en el que los bosones W decaen en τ los cuáles pueden decaer a leptones o hadrones.

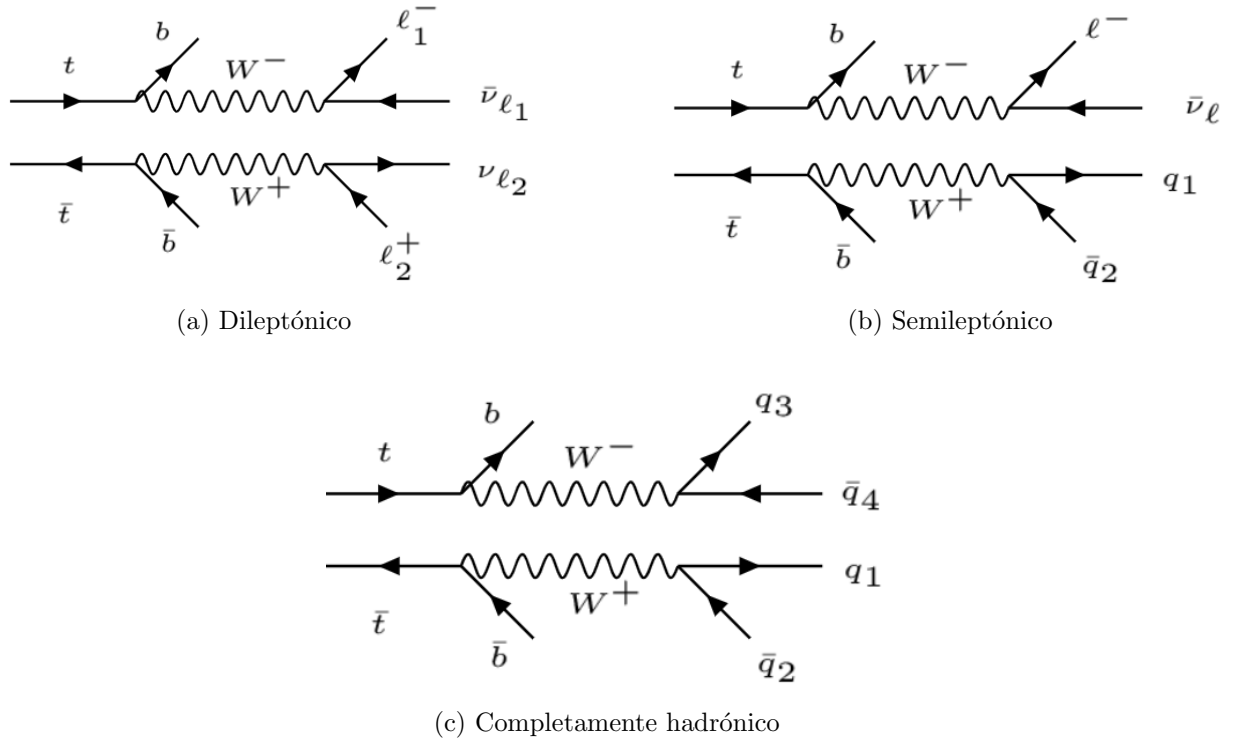


Figura 3: Diagramas de feynman de los distintos modos de decaimiento de los eventos $t\bar{t}$.

A continuación, en la figura 3 se puede observar un diagrama de los posibles decaimientos de los bosones W.

$\bar{c}s$	electron+jets	muon+jets	tau+jets	all-hadronic	
$\bar{u}d$					
τ^-	$e\tau$	$\mu\tau$	$\tau\tau$	tau+jets	
μ^-	$e\mu$	$\mu\mu$	$\mu\tau$	muon+jets	
e^-	$e e$	$e\mu$	$e\tau$	electron+jets	
W decay	e^+	μ^+	τ^+	$u\bar{d}$	14 $c\bar{s}$

Figura 4: Esquema decaimiento W.

2.2.2.1. $t\bar{t}$ en canal semileptónico

El canal semileptónico consiste en que los bosones W que se producen a partir de la pareja top anti-top, uno de ellos decaiga en un leptón (muón o electrón) y el otro de forma hadrónica. De esta forma, en el estado final se encuentra un leptón con alto momento transversal (p_T) y al menos dos b jets que proceden directamente del $t\bar{t}$. El neutrino que se produce en este proceso no se detecta entonces se tendrá una energía restante correspondiente a esta partícula. Este canal se ha escogido porque tiene mucha estadística y por lo tanto los resultados obtenidos serán mucho más precisos. Esta alta estadística se debe principalmente a que las *branching fractions* son muy grandes [10].

2.2.2.2. Fondos en el canal semileptónico

Hay varios procesos, dentro de lo que es el modelo estándar que se puede llegar a confundirse con eventos $t\bar{t}$ en el canal semileptónico. Esto se produce cuando el estado final de ambos decaimientos son iguales o parecidos ya que alguna limitación del sistema de detección u otros factores externos puede llevar a confundirlos con el canal deseado. Con el leptón que aparece en el estado final hay que tener especial cuidado ya que puede venir del vértice de la interacción, los que reciben el nombre de *prompt* pero también pueden ser los que se les denominan *fake*, es decir, que no procedan de la interacción sino que su origen sea otro distinto bien sean los rayos cósmicos, jets mal identificados o que procedan de desintegraciones de mesones. Por lo tanto, se van a considerar los siguientes procesos como fondo:

1. **Dos bosones.** Estos procesos agrupan 3 casos distintos en función de los bosones que intervengan, pueden ser WW, WZ y ZZ. Los tres procesos afectan de una manera muy similar. En el caso de que ambos bosones decaigan de forma hadrónica, en el estado final habrá 4 jets, por lo que si aparece un leptón *fake* y al menos 2 de los jets son b, la señal que producen serán idénticas. Si el bosón W decae de forma leptónica y el otro bosón de forma hadrónica en el estado final hay dos jets, un leptón y un neutrino. En ese caso, la señal es parecida a los eventos $t\bar{t}$ pero faltarían 2 jets los cuáles puede que no se detecten bien debido a por ejemplo a la dirección con la que salen (paralelo eje z). Menos probable que contribuya es el caso en el que el bosón Z decae leptónicamente y el otro bosón de forma hadrónicamente. Aquí el estado final es igual a diferencia que el neutrino es otro leptón por lo que para que sea fondo uno de los leptones no se debe identificar. además no faltaría energía transversal. En el resto de casos es muy difícil que haya contribuciones.

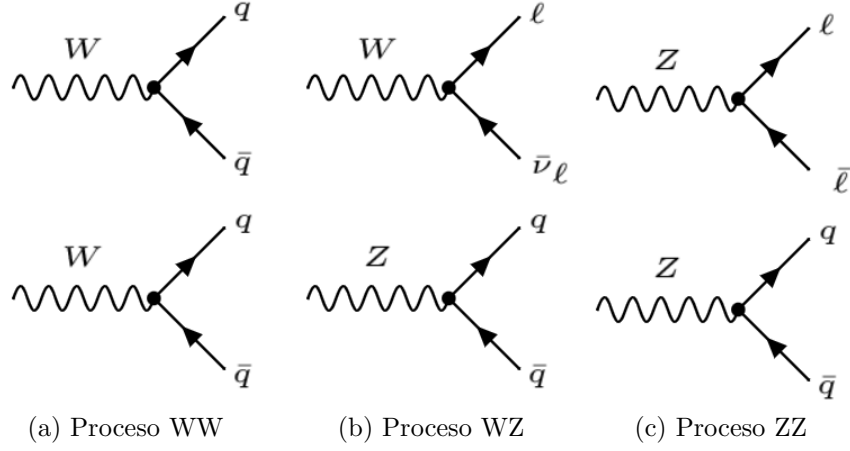


Figura 5: Diagramas de procesos de dos bosones.

2. **$H \rightarrow WW$** Este caso ocurre cuando el bosón de Higgs decae en dos bosones WW lo que al decaer hadrónicamente los dos W o uno de forma leptónica y otro hadrónicamente contribuyen al fondo, tal y como se ha explicado en el caso de los dos bosones.
3. ***Single Top***. Es posible mediante la producción de un único quark top llegar a un estado final con un b jet, otros dos jets, un leptón y energía que falta que correspondería a un neutrino. Esto se produce a partir de un bosón W donde este decae en hadrones y el del vértice tW leptónicamente o viceversa. De esta forma, la única diferencia que hay con los eventos $t\bar{t}$ es un jet, lo que se puede deber a que uno de los jets no se identificó, por ejemplo, por la dirección en la que se produce.

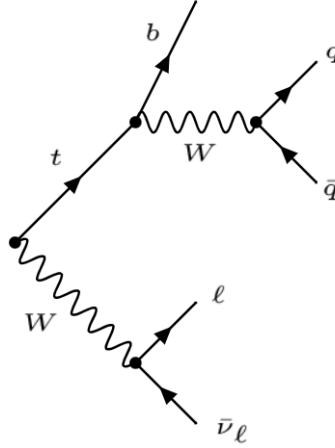


Figura 6: Diagrama para el fondo *Single Top*.

4. **QCD**. La interacción fuerte a través de procesos, como pueden ser la fusión de gluones, pueden llegar a producir un estado final parecido. El leptón que debería aparecer en una señal $t\bar{t}$ podría ser uno mal identificado procedente de un jet (u otras causas mencionadas anteriormente).

Luego se tienen los procesos $t\bar{t}H$ y $t\bar{t}V$ que son propiamente eventos $t\bar{t}$.

- $t\bar{t}H$. Dentro de este proceso se tienen en cuenta dos contribuciones, en función de como decaiga el bosón de Higgs. Si el bosón de Higgs decae en 2 b jets y el $t\bar{t}$ decae de forma semileptónica, la señal sería idéntica a $t\bar{t}$ +jets. En el caso de que $t\bar{t}$ decaiga de forma dileptónica, la señal sería parecida a la de los eventos $t\bar{t}$ pero con un leptón y un neutrino adicional. Por otro lado, también se tienen en cuenta la contribuciones de $t\bar{t}H$ cuando no decae en 2 b jets.

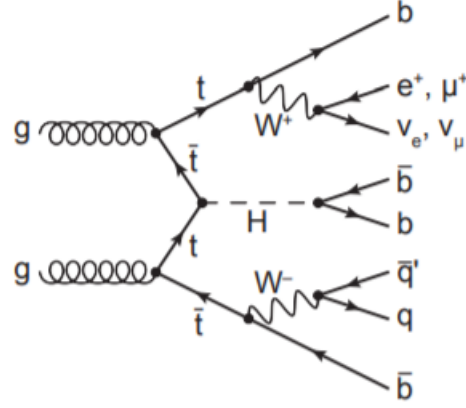


Figura 7: Diagrama del proceso $t\bar{t}H$ [11].

- $t\bar{t}V$. Similar al caso anterior. Si el bosón decae en forma hadrónica ($t\bar{t}$ semileptónico) la señal sería idéntica a la de $t\bar{t}$ +jets.

Por último, para analizar los eventos $t\bar{t}$ con jets adicionales también hay que tener en cuenta los siguientes procesos

- **W+jets**. Este fondo consiste en que un bosón W real producido a su vez con quarks o gluones. El gluón se puede dividir en un par de quarks b produciendo W + jets como Wbb. Como resultado en el estado final habrá 2 b jets, un leptón y energía transversal que falta, lo que es muy parecido a la señal de $t\bar{t}$. Si el W decae hadrónicamente, en el estado final habría 4 jets con dos b jets al menos, por lo que si el leptón identificado fuera un leptón *fake* también sería fondo.

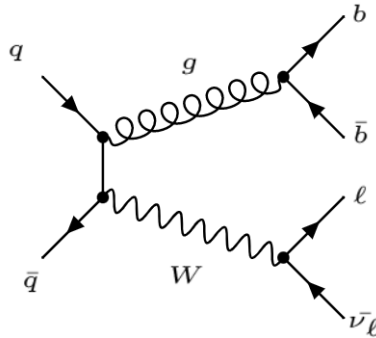


Figura 8: Diagrama de eventos W+jets.

- **Z+jets.** Caso similar al de W+jets pero en este caso el vértice sería Zbb. Aquí si el Z decae leptónicamente habría 2 b jets y 2 leptones, esta señal sería más difícil de confundirla con eventos $t\bar{t}$ pero sería posible que sucediera. En el caso de que el Z decaiga en quark anti-quark se tendría la misma situación que para W+jets.

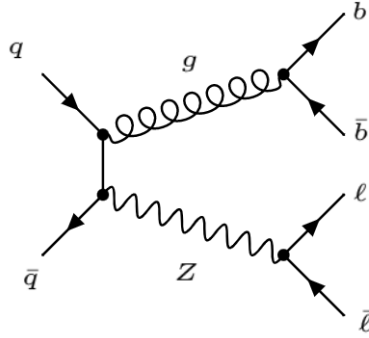


Figura 9: Diagrama correspondiente al fondo Z+jets.

3. Experimento LHC

3.1. Introducción

Los datos que se han analizado provienen del LHC (*Large Hadron Collider*). El LHC, es un acelerador de partículas situado en la frontera entre Francia y Suiza. Consiste en un anillo lleno de imanes superconductores con un perímetro de 27 km. Aquí se producen las colisiones protón-protón (pp) que dan lugar a los procesos que se van a estudiar. Para lograr las colisiones, el método que se sigue es enviar bunches de protones $\approx 10^{11}$, de tal forma que al hacerlos colisionar se aseguran de que siempre se va a producir una colisión con el ángulo correcto. Esta es la razón por la que se envían estos bunches, ya que si sólo se enviarán dos protones esto sería muy difícil de conseguir. La energía de centro de masas que se utiliza en este acelerador ha variado a lo largo de los años. Se han llegado a alcanzar los 13 TeV. Dentro del LHC, hay 4 detectores que se usan en función del propósito del estudio. Uno de estos detectores es ALICE (*A Large Ion Collider Experiment*) el cuál está diseñado para el estudio de la física de la interacción fuerte a unas densidades de energía extremas (plasma quark-gluón). Después también se encuentra el LHCb (*The Large Hadron Collider beauty*) que se encarga del estudio de las diferencias entre materia y antimateria a través del quark b. Los detectores ATLAS (*A Toroidal LHC Apparatus*) y CMS (*Compact Muon Solenoid*) tienen unos propósitos más generales y sus objetivos son los mismos pero está construidos de distintas maneras de tal forma que los resultados obtenidos por uno se puedan comprobar con el otro. Por último, hay otros detectores más pequeños (LHCf, TOTEM, MOEDAL y FASER) que se dedican al estudio de *forward physics* [12].

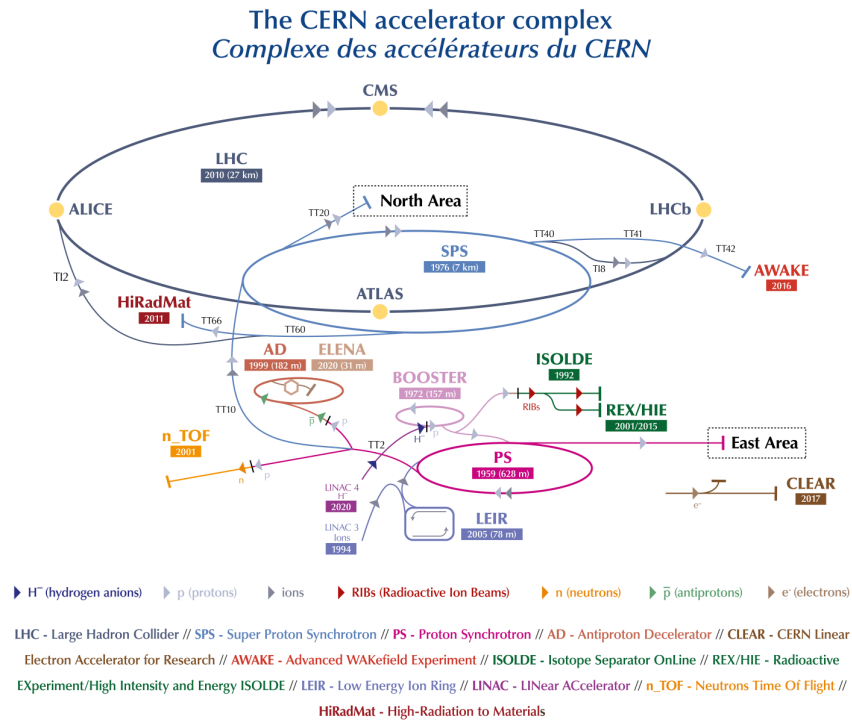


Figura 10: Esquema del LHC y sus detectores.[13]

3.2. CMS

El CMS (*Compact Muon Solenoid*), es un detector del LHC destinado a abarcar una gama amplia de búsquedas y medidas del SM junto al detector ATLAS. De este detector es de donde se han obtenido los datos que han sido analizados. El CMS está construido alrededor de un enorme imán solenoide con forma de bobina cilíndrica de cables superconductores que crean un campo magnético de 3.8 Teslas [14]. Dentro del detector, hay una serie de subdetectores cada uno con una finalidad distinta de forma que se puedan detectar las partículas producidas en las colisiones. A diferencia del ATLAS, su funcionamiento se basa en distintos principios y el sistema magnético es diferente de tal forma que se puedan comprobar los resultados obtenidos en uno de ellos con el otro.

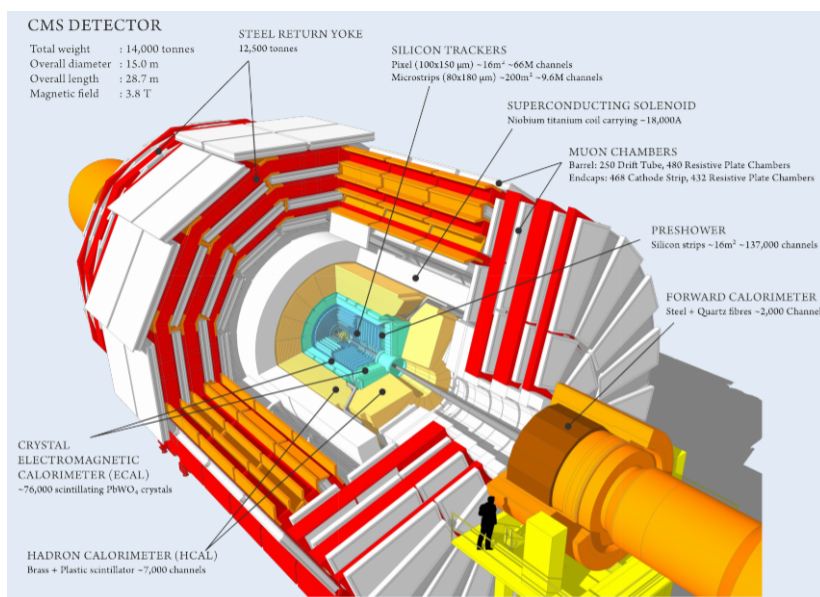


Figura 11: Esquema del CMS, sus dimensiones y subdetectores.[15]

En la figura 11, se puede observar las dimensiones del detector CMS así como los distintos subdetectores que posee para la correcta detección de las partículas. A continuación, en la figura 12, se puede observar como se comportan los distintos tipos de partículas al pasar por cada subdetector.

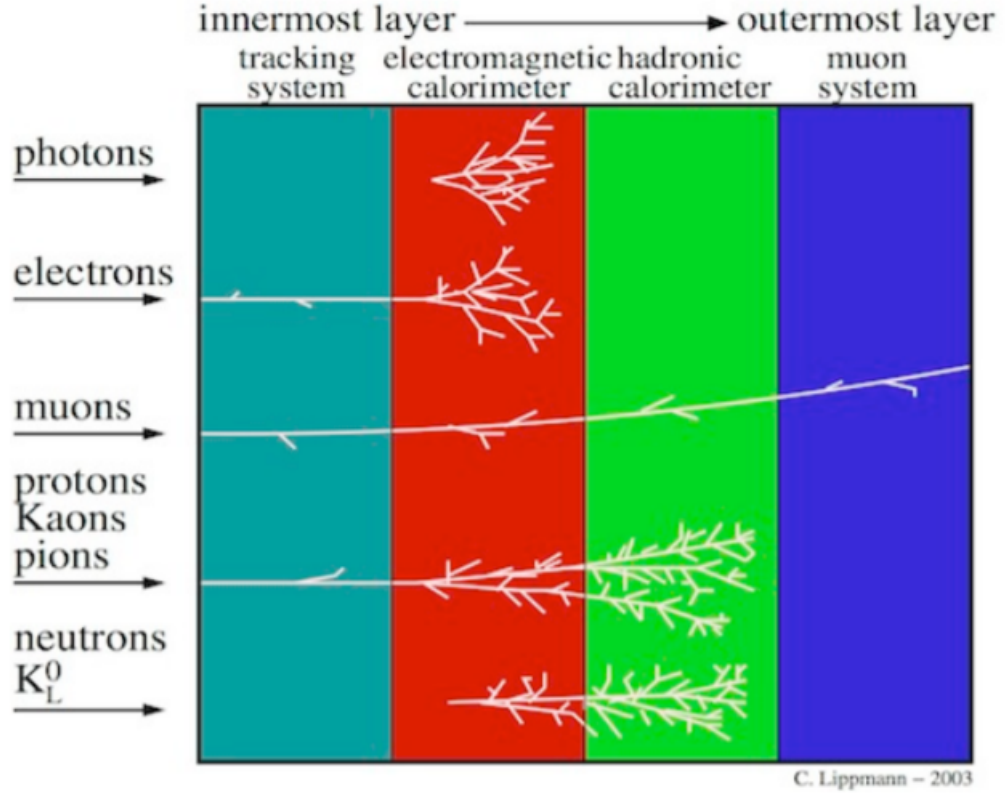


Figura 12: Distintas capas de subdetectores y cómo las distintas partículas interacciona con cada capa.[20]

Por ejemplo, se aprecia como las partículas sin carga eléctrica no es detectada por el *tracking system* o como las partículas electromagnéticas ligeras no superan el *electromagnetic calorimeter*.

3.2.1. Sistema de coordenadas

En el detector CMS se toma como origen de coordenadas el punto de la colisión. A partir de ese punto, en dirección vertical se encuentra el eje y, y en la dirección del haz está el eje z. Por otro lado, el eje x apunta al centro del LHC. Utilizando las coordenadas polares se define el ángulo azimutal (plano XY) como $\phi = \arctan(y/x)$ mientras que el ángulo polar (medido desde el eje z) es $\theta = \arctan\left(\frac{\sqrt{x^2+y^2}}{z}\right)$. Debido a la geometría cilíndrica del detector, en vez de usar las coordenadas esféricas, se ha definido la pseudorapidez, η , para sustituir el ángulo polar θ , cuya relación entre ambos es

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right)$$

De la misma forma, en vez de usar el momento lineal, se utiliza el momento transversal, p_T

que se calcula a partir de la energía transversal depositada en los calorímetros [16].

Este sistema de coordenadas es útil en física de altas energías (HEP) ya que la multiplicidad de partículas de alta energía es aproximadamente constante en η . Además, los intervalos de rapidez son invariantes de Lorentz a lo largo del eje z .

3.2.2. Tracker

El primer detector que se puede observar es el tracker, el cuál es el más cercano al punto donde se produce la colisión. Estos subdetectores usan tecnología basadas en silicio. Su propósito es reconstruir las trayectorias de las partículas cargadas. Para ello, se mide el momento que tienen las partículas que lo atraviesan mediante el uso de campos magnéticos. Las partículas cargadas que atraviesan este subdetector se van a curvar, dependiendo cuanto se curven tendrán más o menos momento (a más curvatura, menos momento y viceversa). Estas trayectorias se reconstruyen a partir de los *hits* que dejan en el detector. Con todo lo explicado de este subdetector, se puede concluir que este detector será sensible únicamente a las partículas con carga eléctrica, como pueden ser los muones o los hadrones cargados.

3.2.3. Calorímetros

Se pueden distinguir dos tipos de calorímetros, electromagnéticos o de hadrones. Estos subdetectores tienen la función de medir la energía depositada por las partículas que lo atraviesan, es decir se absorben las partículas y la energía que dejan es medida.

Los calorímetros electromagnéticos (ECAL) son los subdetectores que van a continuación del tracker encontrándose estos en el interior del solenoide. Aquí, se mide la energía de las cascadas de partículas electromagnéticas ligeras como fotones y electrones o jets, causadas por procesos la producción de pares o el *bremssstrahlung*. El ECAL tiene una buena resolución energética proporcionada por un calorímetro de cristal homogéneo (PbWO_4) [17].

Después se encuentran los calorímetros de hadrones (HCAL), los cuáles rodean al ECAL y se encargan de medir la energía depositada por los hadrones y sus productos de decaimiento a causa de la fuerza fuerte que produce estas cascadas [18]. Estos subdetectores están hechos de capas de cobre intercaladas por capas de un material centelleador.

3.2.4. Detector de muones

El último subdetector y por lo tanto el más externo, es el detector de muones. Se encarga de detectar estas partículas ya que son las únicas, junto a los neutrinos, que no pierden toda o gran parte de su energía en los calorímetros. Este detector está formado por distintas tecnologías de detección. Estas son tres: los *drift tubes* (DTs), los *Cathode Strip Chambers* (CSCs) y los *Resistive Plate Chambers* (RPCs). Su funcionamiento se basa en gas pero de distinta composición.

Los *drift tubes* abarcan el rango de pseudorapidez de $|\eta| < 1,3$ y se encuentra en la zona del barril del CMS. Por otro lado, se tienen los *Cathode Strip Chambers*, los cuáles están situados en las tapas del CMS y detectan muones en el rango $0,9 < |\eta| < 2,4$. Por último,

están los *Resistive Plate Chambers*, los cuáles se encuentran tanto en las tapas como en el barril con una aceptación máxima de $|\eta| \approx 1,6$ y funciona junto a los otros dos haciendo de *trigger*. [19]

En la figura 13, se puede observar un diagrama explicativo de como están distribuidas los distintos subdetectores a lo largo del CMS así como las trayectorias que siguen las distintas partículas.

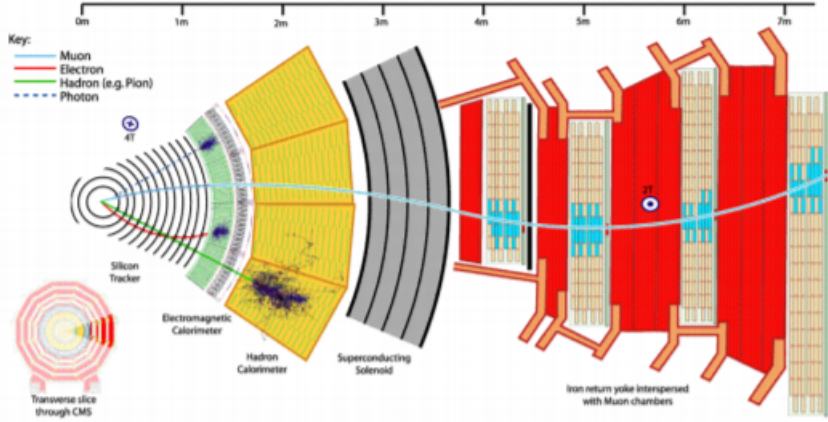


Figura 13: Esquema de las distintas capas y de las trayectorias que siguen las partículas. [21]

Anteriormente se ha explicado como se detectan las distintas partículas que se pueden encontrar, pero en el caso de este detector se puede observar como los neutrinos no son detectados ya que no interactúan con el detector. Pero a pesar de esta situación, haciendo los cálculos se puede ver en ocasiones que para que la conservación de la energía se cumpla hay una parte de la energía transversal que falta. Una de las explicaciones de este fenómeno, es que esta energía pertenezcan a estas partículas que son invisibles para el detector, aunque hay que tener precaución porque también se puede deber esta pérdida de energía a mediciones incorrectas del momento de las partículas detectadas o por efectos de resolución del detector.

3.3. b-tagging

En función del origen de los jets, se distinguen dos tipos: *heavy flavour* (HF) y *light flavour* (LF). Si se originan a partir de quarks b o c se tratan de jets *heavy flavour* pero si producen de quarks u, d o s, son jets *light flavour*. Distinguir estos tipos de jets es muy importante en el estudio del Higgs y también como es este caso, el estudio de la física del top. Esto es debido a que en el estado final de un evento $t\bar{t}$ casi siempre se van a encontrar 2 b-jets.

Por tanto, para distinguir entre ambos tipos de jets hay que atender a las propiedades de estos quarks. Lo más utilizado es el llamado CSV (*combined secondary vertex*) que se basa en el tiempo de vida del quark b. Como los quarks b son partículas con un gran tiempo de vida (1.5ps), se produce un vértice secundario desplazado del punto de decaimiento. Este método es el que se ha utilizado para el *b-tagging*, concretamente en este análisis se ha

usado el algoritmo Deep CSV que se basa en lo explicado anteriormente pero utilizando una *Deep neuronal network* [24] de tal forma que mejora la eficiencia del etiquetado de jets convencional.

Para ello, a cada jet se le asigna un número de 0 a 1 en función de la probabilidad de que sea o no sea un b jet y cada punto de trabajo (WP) tendrá un valor de corte que distinga entre ambos tipos de jets. Dentro de este algoritmo, hay distintos puntos de trabajo en función de su eficiencia. Estos pueden ser *Loose*, *Medium* y *Tight* y se definen como los valores de corte para los cuales la tasa de identificación errónea de un jet LF como un b jet es del 10 %, 1 % y 0,1 %, respectivamente [23].

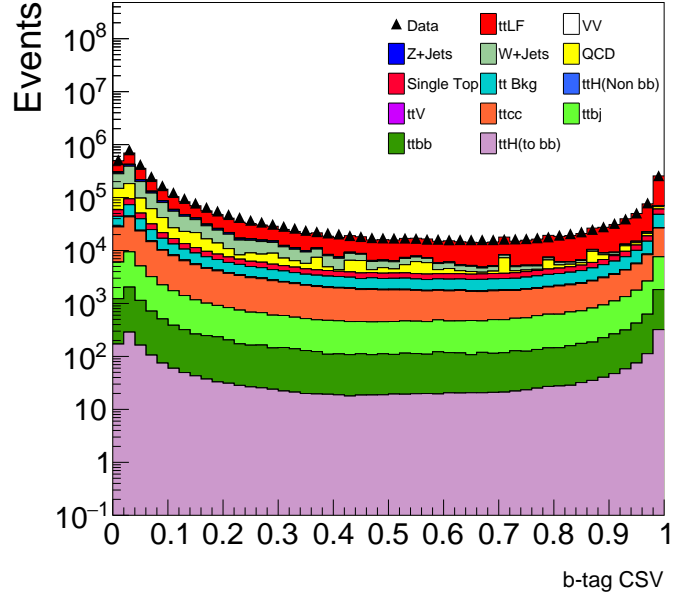


Figura 14: Distribución de b-tag CSV para la región ≥ 4 jets, en escala logarítmica

4. Análisis

El canal de decaimiento del par top anti-top que se ha estudiado en este trabajo es el semileptónico, es decir, de los dos bosones W que se producen en el decaimiento, uno decae en leptones (muón o electrón) y el otro de forma hadrónica. Para registrar estos eventos se utilizan *triggers* que consisten en una serie de criterios para almacenar los eventos que son relevantes para el estudio que se va a realizar. En este caso, como el canal de estudio es el semileptónico, se ha utilizado un *Single lepton (μ/e) trigger* de tal forma que solo se seleccionen los eventos con un único leptón (muon o electrón). A la hora de seleccionar estos eventos, se imponen una serie de criterios adicionales como son la calidad, el momento transversal y la pseudorapidez. Por otro lado, en el canal semileptónico del decaimiento del par top anti-top, además de dos bosones W también se producen dos quarks b, los cuales dan lugar a dos b jets. Por lo tanto, este es otro criterio que se ha utilizado para la selección de los eventos, es decir, se impone que al menos haya dos quarks b y además el número de jets sea mayor o igual que 4. Esto se debe a que en el caso del decaimiento semileptónico uno de los bosones W va a decaer de forma hadrónica, típicamente en dos *light flavour jet*.

4.1. Objetos

A continuación, se han explicado los criterios de selección mencionados anteriormente para la selección de objetos.

Para realizar una correcta identificación de los electrones y muones, además de utilizar los criterios usados por la colaboración del detector, se ha exigido que el pT de los electrones sea mayor de 35 GeV y en el caso de los muones el momento transversal tiene que superar los 30 GeV. Al exigir un momento transversal tan alto, se reduce la contribución de otros procesos con leptones del decaimiento de los quarks *bottom/charm*, la identificación de un electrón aislado cuando en realidad forma parte de un jets o casos similares. También se ha tomado que la máxima aceptación es $|\eta| \leq 2.4$, lo que es lógico porque en el detector de muones, el máximo $|\eta|$ para el que puede detectar muones es ese valor. La colaboración CMS tiene sus propios criterios de selección de identificación (ID) y de aislamiento (ISO). En este caso, se han utilizado los criterios mas restrictivos, (*tight*), en ambos casos. Esto permite tener una eficiencia superior al 90 %.

Estos leptones que se quieren detectar tienen que estar aislados, es decir, no pertenecer a un jet. Los jets son el conjunto de partículas creadas en la hadronización y por tanto tiene un mismo origen. Como estos jets pueden contener muones y estos pueden confundirse con los muones producidos por el modo de desintegración que se está estudiando. Para ello, se utiliza el criterio de aislamiento previamente mencionado.

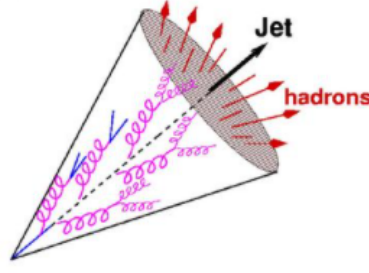


Figura 15: Ilustración de un jet.[22]

Las partículas que forman parte de un jet están contenidas en un cono cuyo radio de la base, ΔR , se ha define como

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \quad (1)$$

Lo más habitual es considerar que los jets tienen ΔR entre 0.3 y 0.5 unidades. En este caso, se ha tomado que el radio tiene un tamaño de 0.4 unidades, de tal forma que las partículas contenidas en el cono de ese tamaño sean identificadas como un jet y no como partículas aisladas. Al igual que los leptones, se han utilizado los criterios más restrictivos para seleccionar los jets. Más adelante, en función de las regiones que se hayan querido estudiar, se han delimitado los eventos estudiados en función del número de jets y su clasificación.

A la hora de identificar los b-jets, se ha utilizado el criterio ya explicado de *b-tagging* basado en el *Combined secondary vertex*. A nivel de análisis, se van a evaluar diferentes cortes viendo la relación señal fondo obtenida, para así determinar cuál es más eficaz, de tal forma que el corte sea muy puro pero que no se pierdan demasiados eventos. En los eventos $t\bar{t}$ se impone el número de b jets en función de las características del canal explicadas anteriormente.

4.2. Categorización de eventos $t\bar{t}$ + jets

Los eventos $t\bar{t}$ pueden producirse con jets adicionales (j) los cuáles pueden ser producidos a partir de quarks de cualquier sabor (salvo t) o gluones. De esta forma, se estudian los pares $t\bar{t}$ junto al a un par de jets (por lo menos). Este proceso se denota como $t\bar{t}jj$ y en simulación han sido categorizados usando el sabor de las partículas de los jets encontradas en la información del generador MonteCarlo. Es identificado mediante la técnica *ghost-matching*. Se distinguen cuatro categorías en función del sabor de los jets que acompañan al par $t\bar{t}$:

- El proceso $t\bar{t}b\bar{b}$ es aquel en que los dos jets adicionales son b-jets.
- El proceso $t\bar{t}c\bar{c}$ consiste en que ambos jets adicionales son c-jets.
- El proceso $t\bar{t}bj$ se da cuando solo uno de los jets adicionales es un b jet.
- El proceso $t\bar{t}LF$ contiene todo los eventos en los que los jets adicionales son *light flavour* y también incluye los eventos sin jets adicionales

El proceso $t\bar{t}b\bar{b}$ es el fondo dominante del estudio del proceso $t\bar{t}H$.

4.3. Gráficos de control

El análisis de los datos se va a centrar en tres procesos concretos $t\bar{t}$, $t\bar{t}b\bar{b}$ y $t\bar{t}H$ para lo cual es necesario establecer unas regiones de estudios. Estas regiones se han creado atendiendo a las propiedades de cada proceso, para obtener el mayor número de eventos posibles. En el caso del proceso $t\bar{t}$, se ha escogido la región de al menos 4 jets donde dos de ellos (al menos) son b-jets. Estos se debe a que en el canal semileptónico, el estado final son dos b jets y luego los bosones W uno decae en un leptón y un neutrino y el otro en dos jets que pueden ser de cualquier tipo. En cuanto a los procesos $t\bar{t}b\bar{b}$ y $t\bar{t}H$, la región que se ha considerado más apropiada es la de al menos 6 jets, 4 de los cuáles, al menos son b-jets. El razonamiento es igual al caso anterior, ya que el proceso puramente $t\bar{t}$ tiene un estado final de al menos 4j y al menos 2 b-jets, por lo tanto en el proceso $t\bar{t}b\bar{b}$ se añaden al menos los dos b-jets adicionales y en el proceso $t\bar{t}H$ como el bosón de Higgs decae en dos b-jets el estado final es el mismo que el proceso anterior. Por último, también se han utilizado las regiones de al menos 4 jets y al menos 6 jets que permiten ver los datos con una perspectiva más general y son útiles a la hora de la creación de las nuevas variables.

Lo primero que se ha llevado a cabo es representar unas variables estándar (en las distintas regiones) para comprobar que la simulación de MonteCarlo es correcta y así validar esta simulación.

4.3.1. Regiones de ≥ 4 jets y ≥ 6 jets

A continuación se muestran las distribuciones de la multiplicidad de los b-jets para las regiones ≥ 4 jets y ≥ 6 jets.

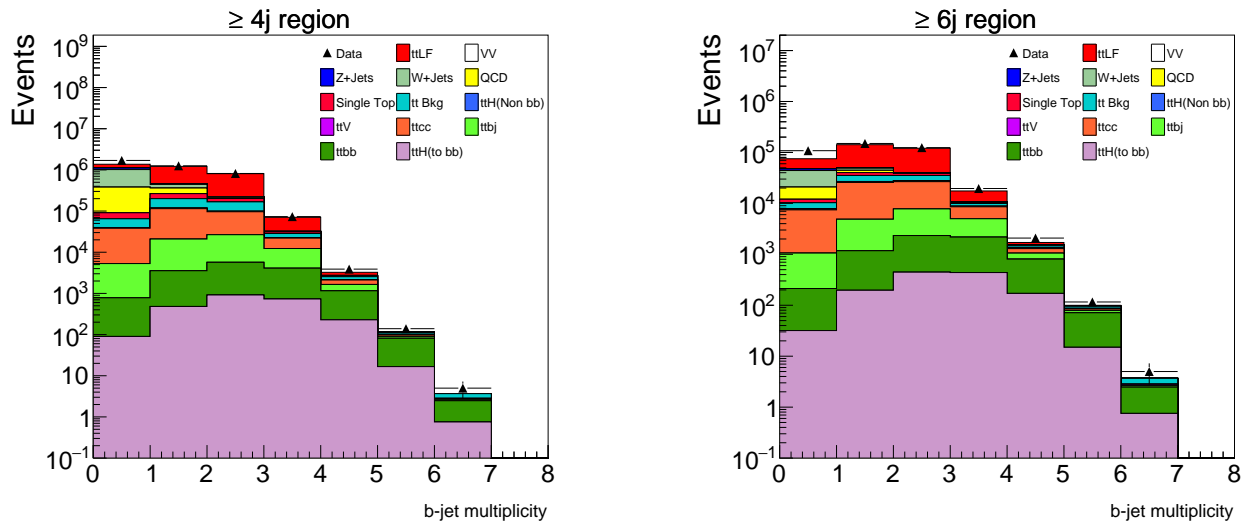


Figura 16: Multiplicidad de los b-jets para las regiones ≥ 4 jets y ≥ 6 jets.

En la figura 18, se puede observar la gráfica correspondiente a la multiplicidad de los b-jets para dos regiones distintas. En la leyenda se puede observar que color corresponde a cada proceso y los triángulos son los datos recogidos. En la izquierda se ha representado la región ≥ 4 jets y en la derecha ≥ 6 jets. Ambas representaciones tienen distribuciones similares

aunque no iguales. Un ejemplo de esto se puede apreciar en el número de eventos de QCD, ya que para la contribución para la región ≥ 4 jets es considerablemente mayor que en la región de ≥ 6 jets. Otra cosa interesante que se puede observar es cómo a partir de los dos b-jets los procesos QCD y W+jets prácticamente desaparecen por lo que el fondo disminuiría.

4.3.2. Región para eventos $t\bar{t}$

Para esta región, en la que se impone que al menos haya 4 jets y 2 b-jets, sean representado otras variables como pueden ser la pseudorapidez o el número de jets. Además se han representado para distintos puntos de trabajo (WP). En estos gráficos los eventos $t\bar{t}$ sería la suma de $t\bar{t}b\bar{b}$, $t\bar{t}b\bar{b}$, $t\bar{t}b\bar{b}$ y $t\bar{t}LF$.

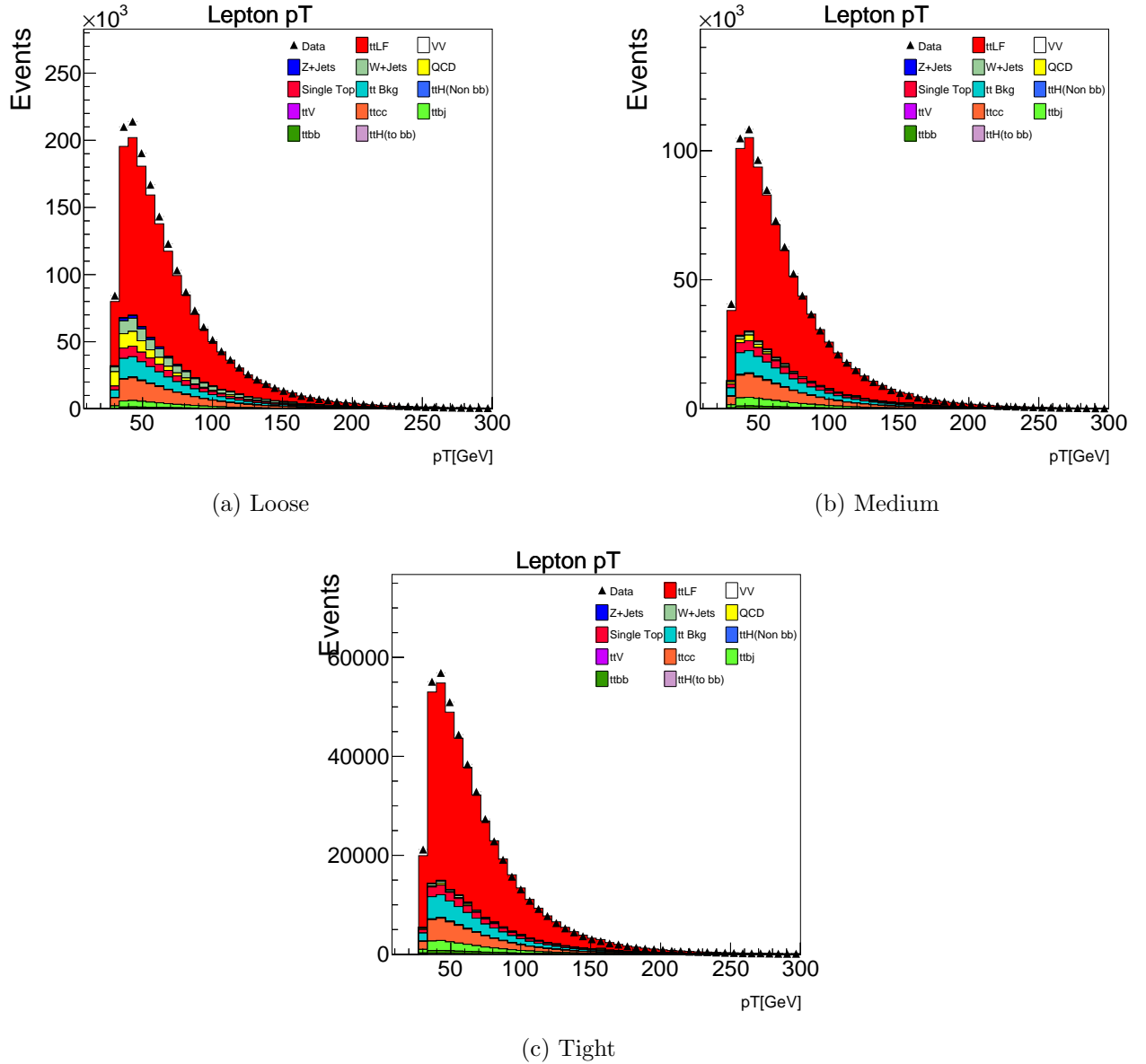


Figura 17: Momento transversal del leptón en GeV para los tres puntos de trabajo en la región de al menos 4 jets donde dos de ellos, por lo menos, tienen que ser b-jets.

En la figura 17, se ha representado el momento transverso del leptón para los tres puntos de trabajo. Aquí se puede observar el corte realizado en los 30 y 35 GeV en función del tipo de leptón. También la forma de la distribución es la esperada.

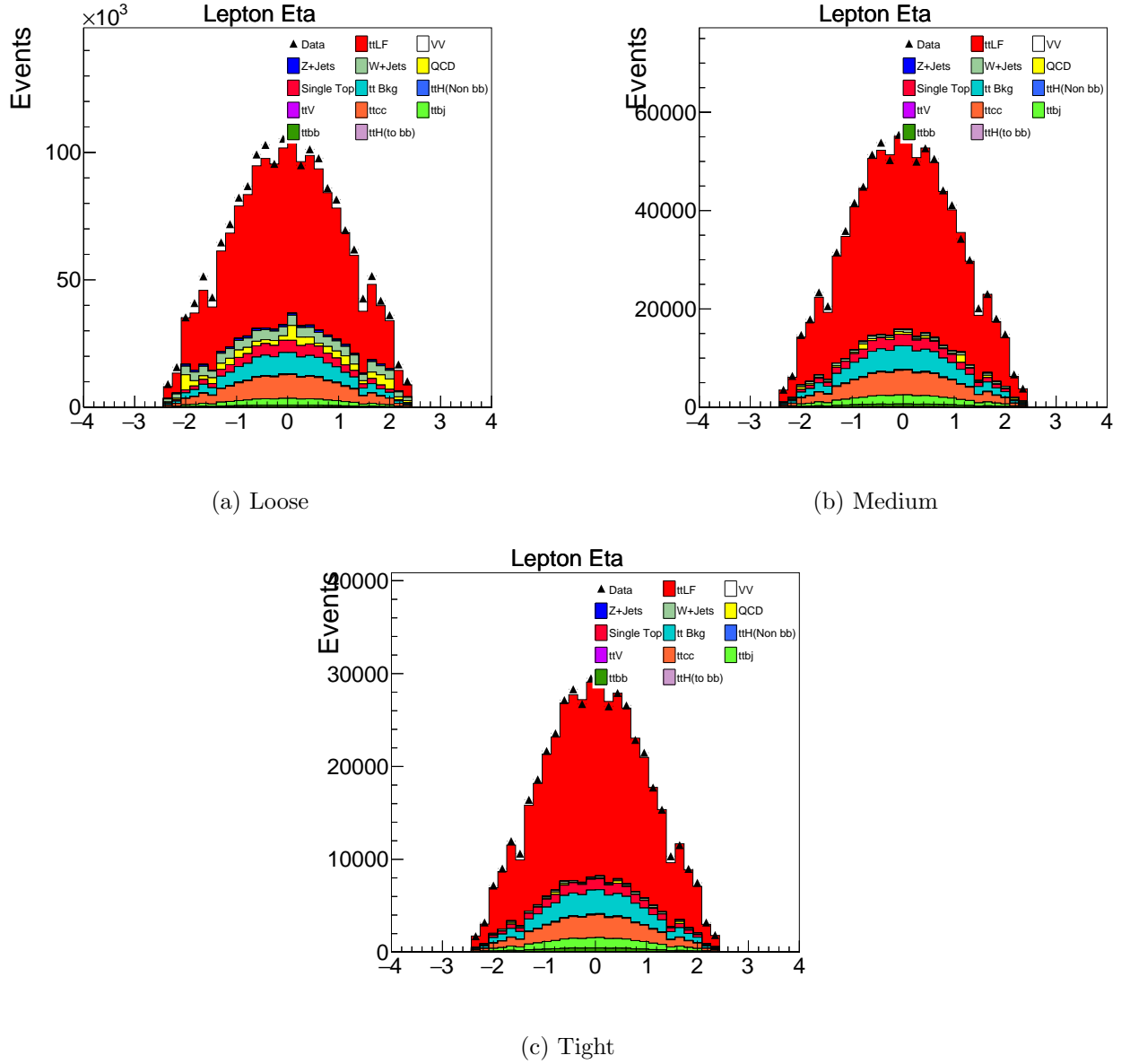
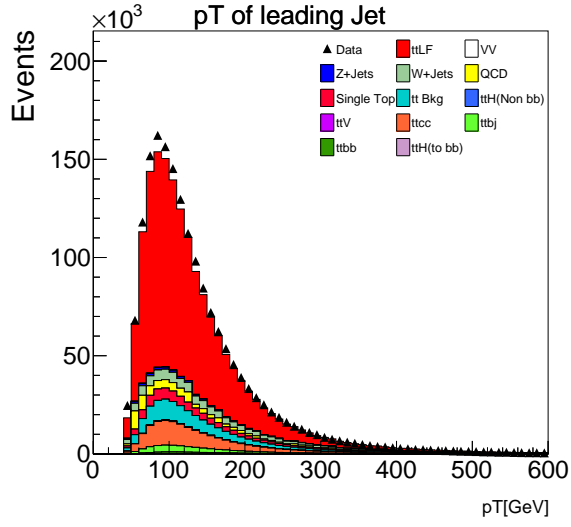
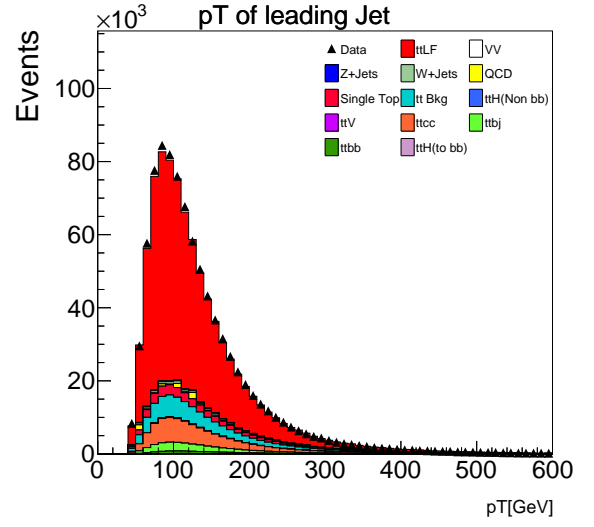


Figura 18: Pseudorapidez de los leptones para los tres puntos de trabajo en la región de al menos 4 jets donde dos de ellos, por lo menos, tienen que ser b-jets.

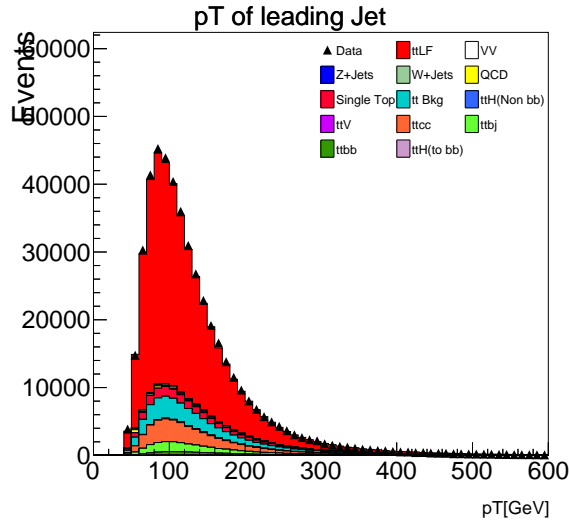
También se obtiene lo esperado en la figura 18. Donde se puede apreciar una distribución simétrica de los datos y el corte coincide con el establecido en los criterios de selección $|\eta| \leq 2,4$.



(a) Loose



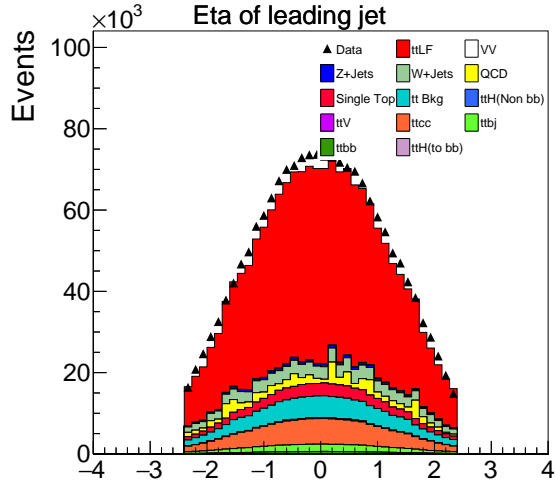
(b) Medium



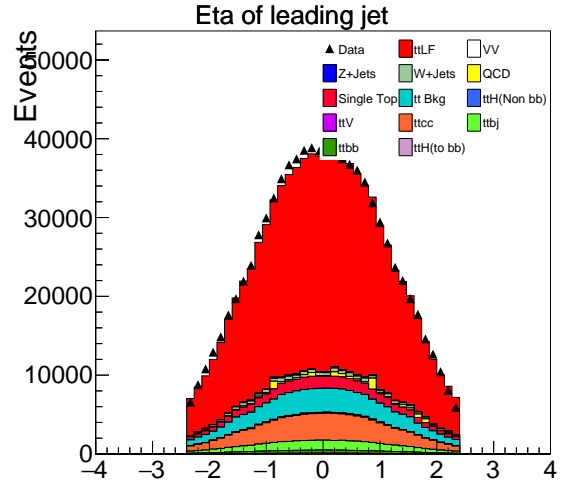
(c) Tight

Figura 19: Momento transversal del primer jet (en GeV) para los tres puntos de trabajo en la región de al menos 4 jets donde dos de ellos, por lo menos, tienen que ser b-jets.

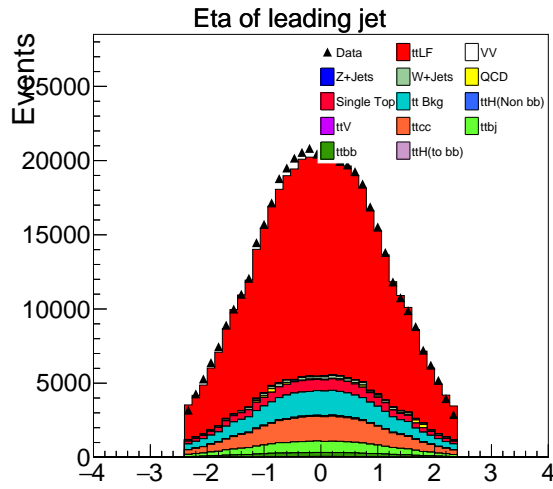
Se puede observar como la figura 19 tiene la forma característica del momento transversal.



(a) Loose



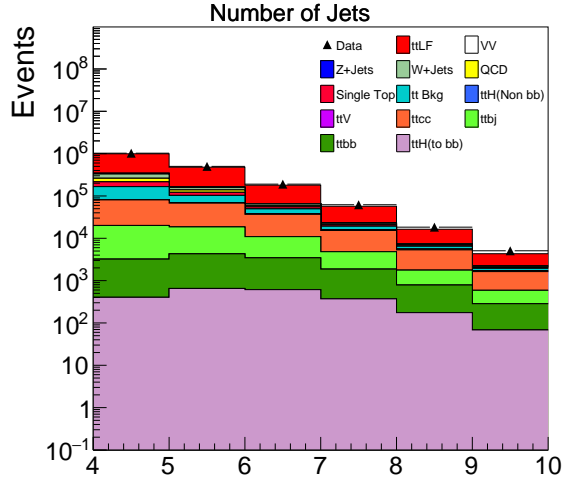
(b) Medium



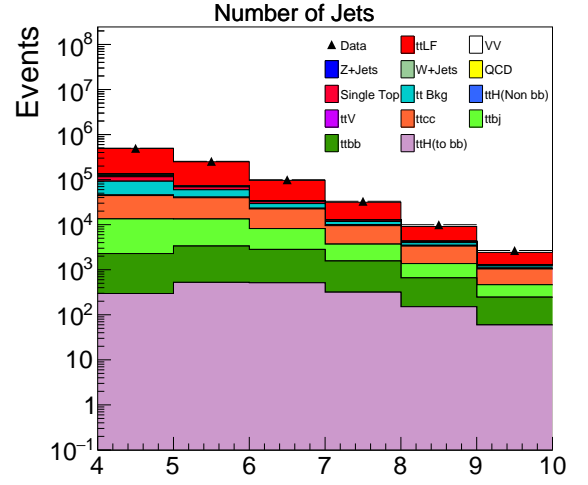
(c) Tight

Figura 20: Pseudorapidez del primer jet en para los tres puntos de trabajo en la región de al menos 4 jets donde dos de ellos, por lo menos, tienen que ser b jets.

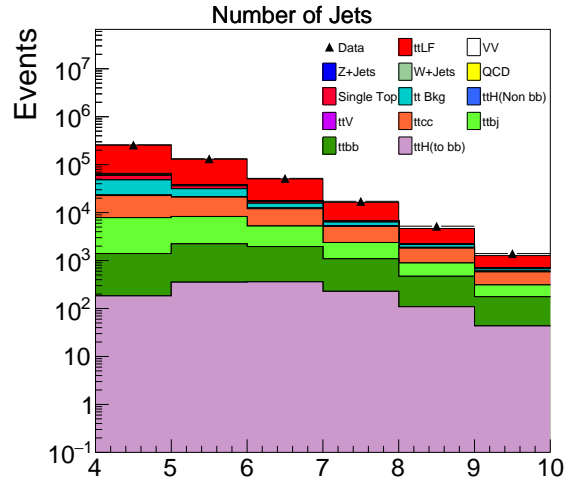
En la figura 20 al igual que la figura 18 tiene la característica distribución simétrica. También se puede observar el corte aplicado claramente.



(a) Loose



(b) Medium



(c) Tight

Figura 21: Número de jets en los tres puntos de trabajo para la región de al menos 4 jets de los cuáles al menos dos son b-jets en escala logarítmica.

En la figura 21 se puede ver como está distribuido el número de jets en la región y punto de trabajo seleccionado. Los resultados son los esperados ya que conociendo el decaimiento que se está estudiando se esperaba que la mayoría de los procesos tuvieran entre 4 y 6 jets.

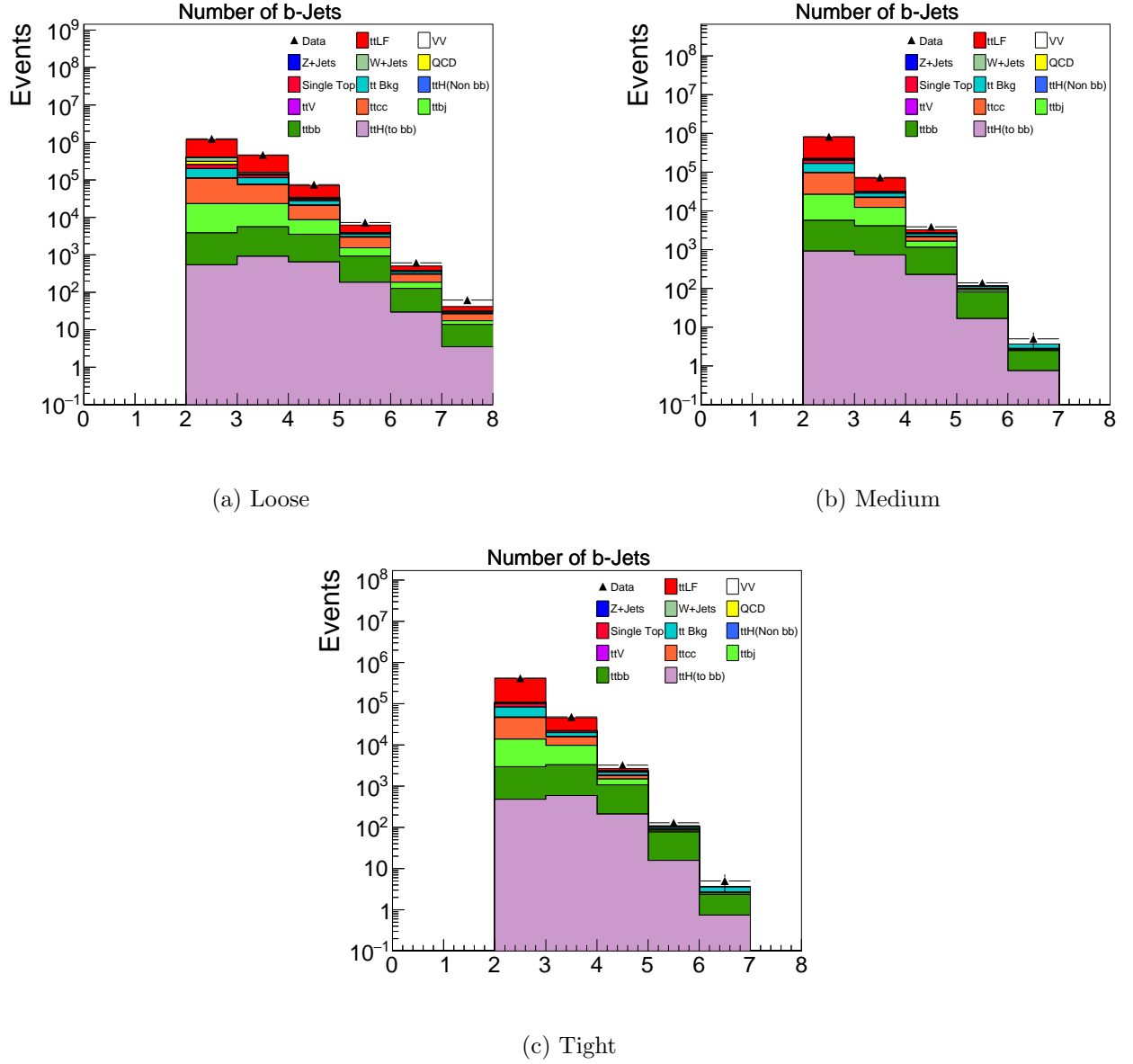


Figura 22: Número de b-jets en los tres puntos de trabajo para la región de al menos 4 jets de los cuáles al menos dos son b-jets, en escala logarítmica.

En la figura 22 se puede ver como está distribuido el número de b-jets en la región y punto de trabajo seleccionado. Los resultados son los esperados ya que conociendo el decaimiento que se está estudiando se esperaba que la mayoría de los procesos tuvieran alrededor de 2 b-jets.

En general, también se puede ver como utilizando el punto de trabajo Loose hay una mayor contribución del fondo que en los otros dos casos. Esto se debe a que es el WP menos restrictivo de los tres existentes y por lo tanto, aumenta la probabilidad de que jets LF se confundan con b-jets. Muestra de esto es el aumento de eventos correspondientes a los procesos W+jets y QCD cuando el WP es Loose, respecto a los otros dos casos, lo que se

puede apreciar en la figura 22.

4.3.3. Regiones para eventos $t\bar{t}$ +jets

Para analizar estos procesos, se han utilizado dos regiones concretas. La primera de ellas, la más general, corresponde con al menos 6 jets donde 2 de los cuáles, por lo menos, tienen que ser b. Esta región se ha utilizado para estudiar, en general, los casos $t\bar{t}$ +jets ya que en el estado final se tendrían los 2 jets adicionales más los 4 jets, donde dos de los cuáles son b propios de los eventos $t\bar{t}$ (canal semileptónico). Por último, para estudiar específicamente el proceso $t\bar{t}b\bar{b}$ se define la región de al menos 6 jets donde 4 son b jets. Esto viene al igual que antes porque en el estado final de los $t\bar{t}$ semileptónicos son 4 jets donde 2 de los son b y ahora los dos jets adicionales son b jets. Esta misma región se utiliza para el proceso $t\bar{t}H$, donde el bosón de Higgs decae a dos b-jets.

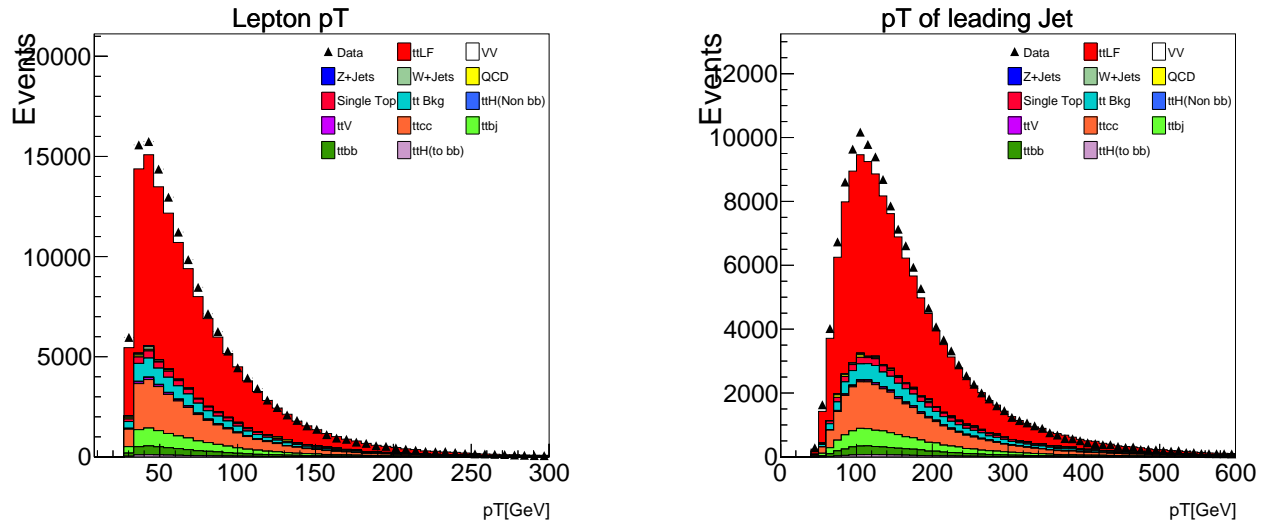


Figura 23: Momento transversal del leptón y del primer Jet en GeV para el punto de trabajo *Medium* en la región de al menos 6 jets donde dos de ellos, por lo menos, tienen que ser b-jets.

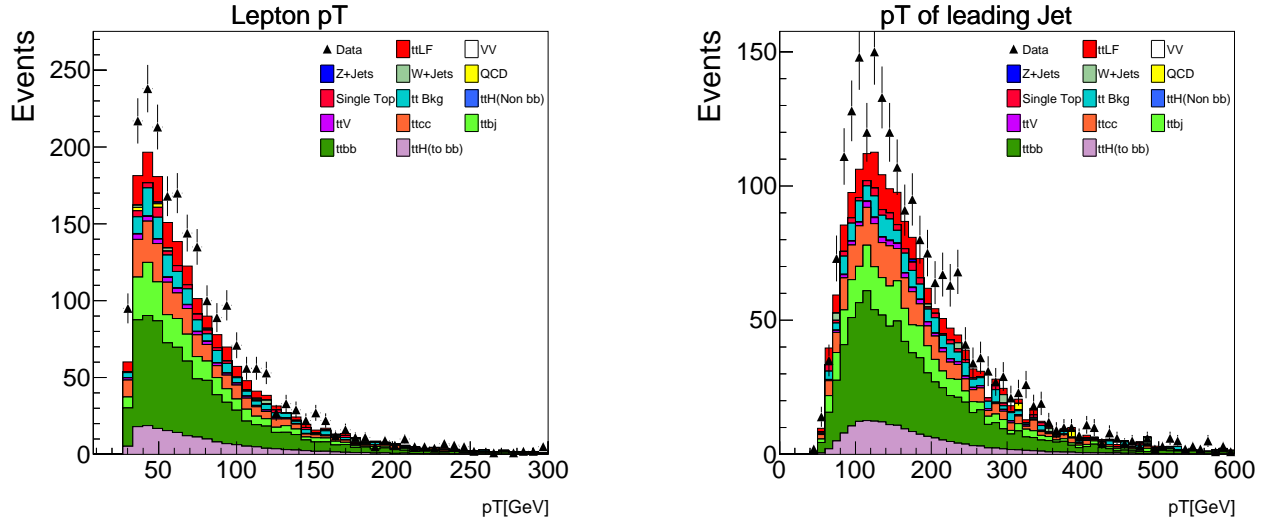


Figura 24: Momento transversal del leptón y del primer Jet en GeV para el punto de trabajo *Medium* en la región de al menos 6 jets donde cuatro de ellos, por lo menos, tienen que ser b-jets.

4.4. Construcción de variables para la identificación de eventos con b-jets

A lo largo del proyecto se ha explicado que es fundamental distinguir los tipos de jets que se detectan, es decir, conocer si son *light flavour* o *heavy flavour* jets. Para ello, se ha utilizado el método DeepCSV basado en que los quarks b tienen un gran tiempo de vida lo que produce un vértice secundario desplazado del punto de decaimiento. El problema de este método es que es demasiado restrictivo lo que puede hacer que potenciales b jets no se identifiquen correctamente. Para solucionar este problema, se han creado dos variables, la primera corresponde con la suma de los discriminantes del CSV de cada uno de los jets

$$\sum_{i=0}^{n_{jets}} CSV_{jet_i}$$

De esta forma, se puede analizar la cantidad de b jets desde la perspectiva del evento que se está estudiando y no individualmente. Siguiendo en esta línea se ha definido la segunda variable que recibe el nombre de sumatorio relativo de b-tags

$$\frac{\sum_{i=0}^{n_{jets}} CSV_{jet_i}}{\#jets}$$

donde $\#jets$ hace referencia la número de jets. Básicamente consiste en hacer una media de todos los b-tags de los jets del proceso. Estas dos variables, ofrecen una perspectiva distinta de la identificación de los jets y menos drástica que la del etiquetado de jets explicado anteriormente. Esto se debe a que por ejemplo si se tiene un evento con dos jets con btags de 0.7 y 0.49 respectivamente según el punto de trabajo medio sólo habría un b jet ya que

el segundo no se consideraría un b-jet por solamente 0.01. En cambio con estas variables, si el proceso predice que hay 2 b-jets, no descartaría que ambos pudieran ser b jets. Por lo que estas variables, podrían mejorar esta identificación para así obtener una mayor pureza y eficiencia.

4.5. Distribuciones de ≥ 4 jets y ≥ 6 jets para las nuevas variables

Una vez que se han creado las dos variables, se han representado en las representado en distintas regiones. La región de al menos 4 jets es la correspondiente al estudio de los eventos $t\bar{t}$ debido al estado final de este proceso. De igual manera, se tiene que la región de al menos 6 jets corresponde a los procesos $t\bar{t}b\bar{b}$ y $t\bar{t}H$.

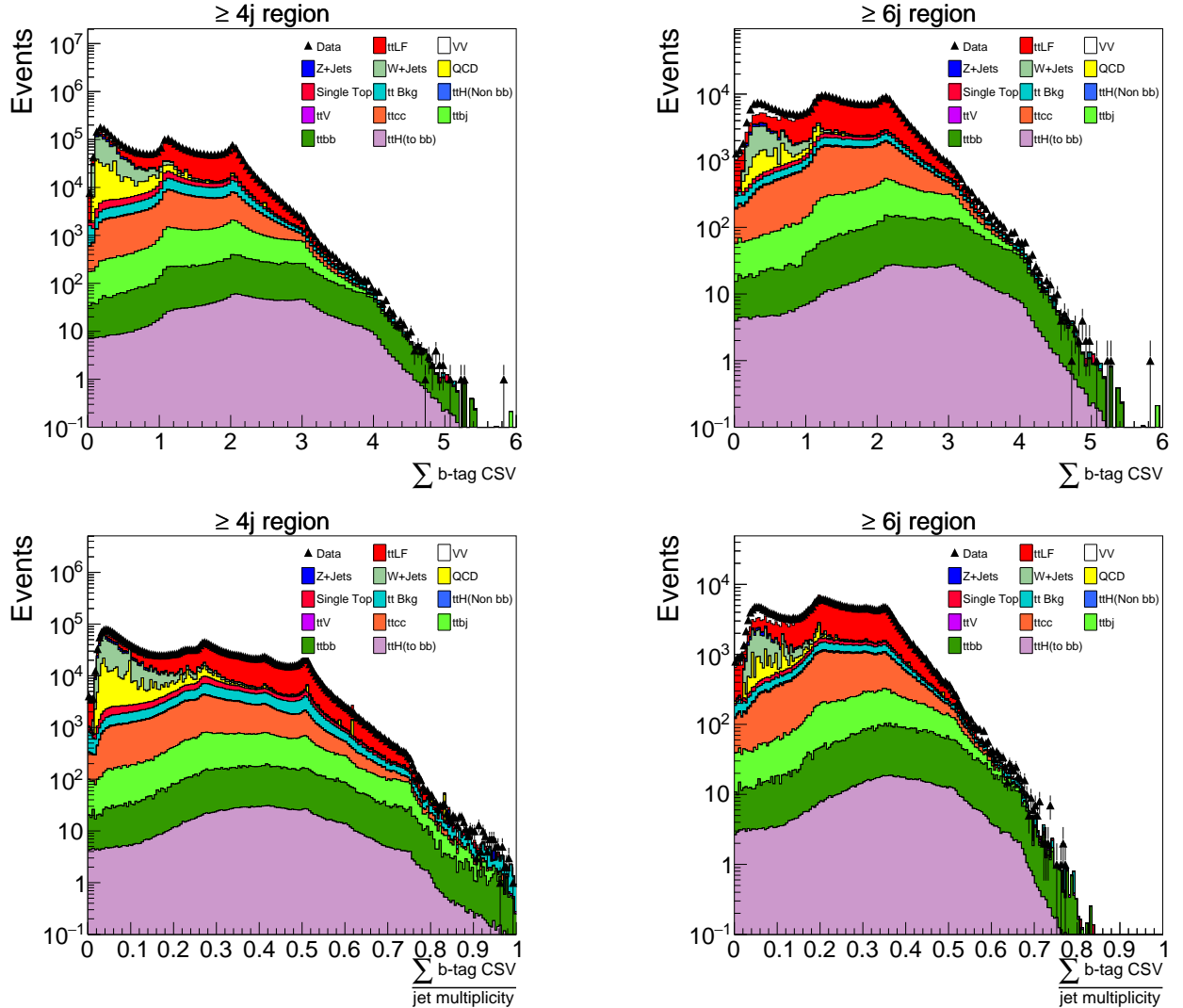


Figura 25: Representación del sumatorio de los b-tags y el sumatorio relativo de b-tags, para las regiones ≥ 4 jets y ≥ 6 .

En ambas figuras, se puede observar como las contribuciones de cada proceso va variando en función de los distintos valores de las variables sumatorio b-tag y el sumatorio relativo

de b-tags. A valores bajos de estas variables, se aprecia como hay muchas contribuciones en especial de los eventos $t\bar{t}LF$, así como otros procesos que actúan como fondo de la señal que se quiere estudiar ($t\bar{t}b\bar{b}$). A medida que los valores aumentan estas contribuciones se hacen más pequeñas y los procesos $t\bar{t}b\bar{b}$ y $t\bar{t}H$ son los principales, con unos eventos de fondo muy reducidos ya que en sus estados finales se encuentran al menos 4 b-jets (más que en los otros procesos).

4.6. Eficiencia y significancia

Vistas la figura 25, se van a establecer una serie de cortes en las dos variables construidas para así elegir la región en la que los resultados equilibren el número de eventos rechazados de la relación señal y fondo. Cómo ya se observa en esa figura, se puede apreciar como a medida que aumenta el valor de estas variables, los procesos $t\bar{t}b\bar{b}$ y $t\bar{t}H$ contribuyen más que el resto. Por lo tanto, en vez de establecer de una forma intuitiva estos cortes, se va a determinar cuáles son los mejores posibles. Para ello, se define la eficiencia como la razón de la cantidad de eventos de la señal que se desea estudiar entre en el número total de eventos. Por otro lado, la significancia es

$$Significancia = \frac{S}{\sqrt{S+B}}$$

Donde S son los eventos correspondientes a la señal y B los referidos al fondo. Se han representado estas variables para determinar cuál es el corte óptimo que se debe realizar. Esto ocurre cuando la significancia es máxima. A continuación se han representado las eficiencias de la señal y el fondo, así como su significancia para las tres variables. Al igual que antes. para la región ≥ 4 , se estudia los eventos $t\bar{t}$ y para la región ≥ 6 los $t\bar{t}b\bar{b}$.

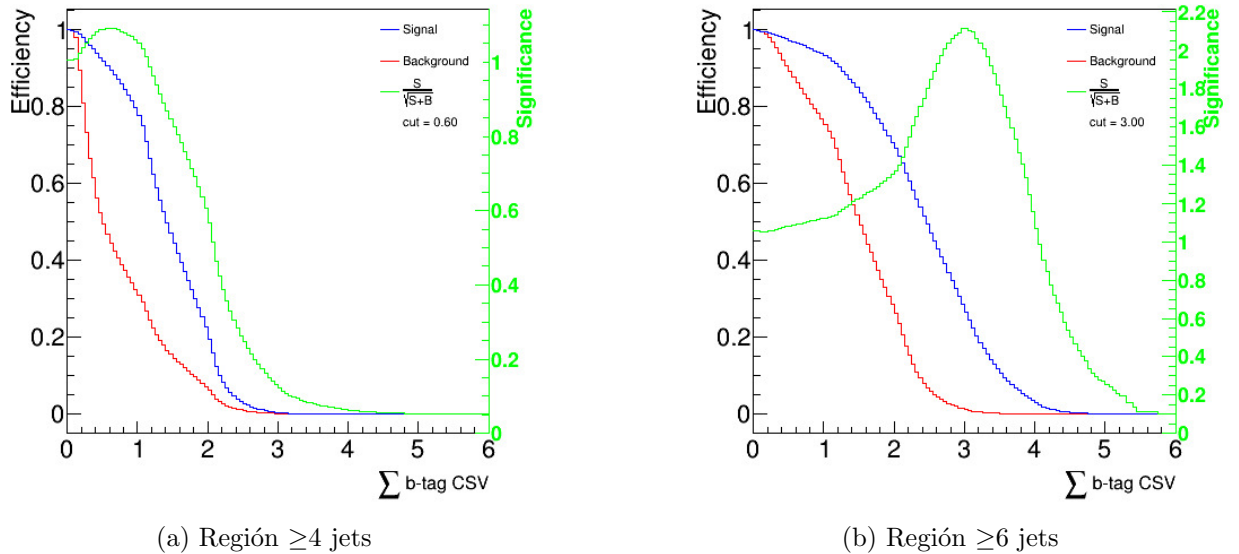


Figura 26: Representación de la eficiencia de la señal y del fondo, así como la significancia, de la variable $\sum_{i=0}^{n_{jets}} CSV_{jet_i}$. En la leyenda aparece *Signal* que es el proceso que se quiere estudiar y *Background* que es el fondo.

En la figura 26 se puede observar como el mejor corte para la región de al menos 4 jets es 0.6, mientras que para la de al menos 6 jets es cuando el sumatorio de b-tags vale 3. Estos números corresponden con el máximo de la significancia representada en verde. Aquí ya se puede apreciar la importancia de la significancia, ya que analizando sólo las eficiencias podría parecer que el mejor corte se encuentra en valores menores donde la eficiencia de la señal parece mucho mayor que la de fondo.

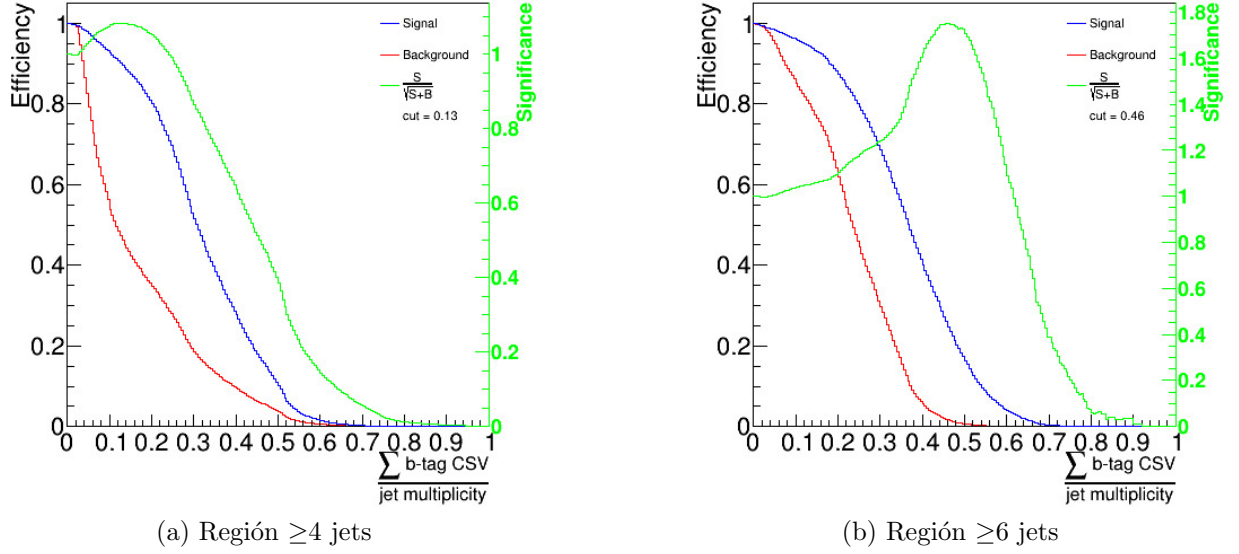


Figura 27: Representación de la eficiencia de la señal y el fondo, así como la significancia, de la variable $\frac{\sum_{i=0}^{n_{jets}} CSV_{jet_i}}{\#jets}$.

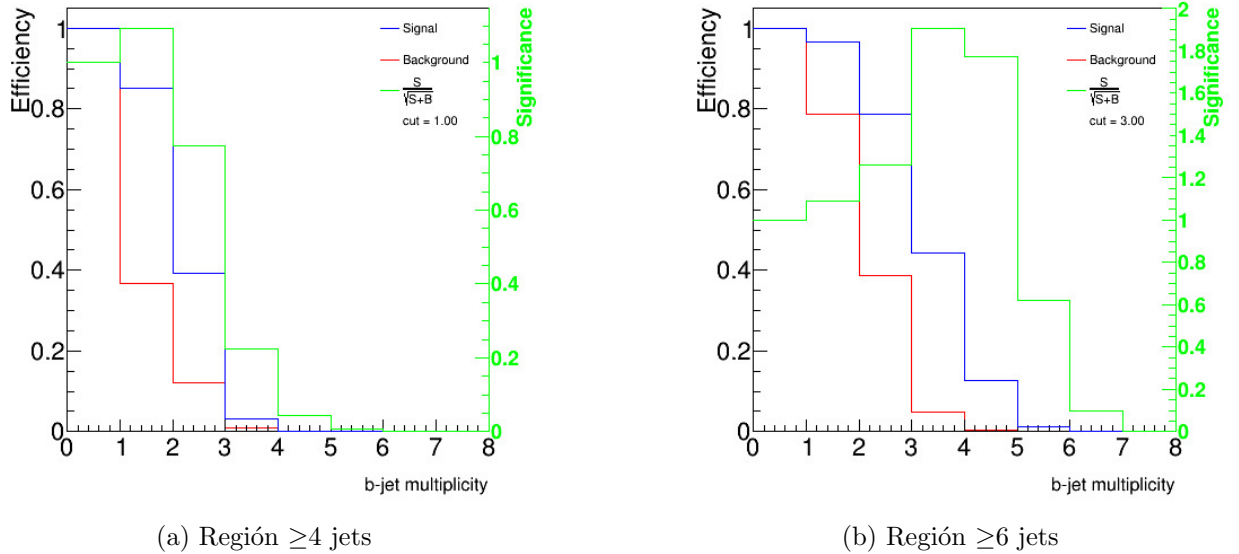


Figura 28: Representación de la eficiencia de la señal y el fondo, así como la significancia, de la multiplicidad de b-jets.

En la figura 27, se ha representado la variable de la suma relativa de los b-tags. Se puede observar, al igual que en el caso anterior, que como es lógico el máximo de la significancia se encuentra a menores valores de la suma relativa que para 6j. De esta manera los cortes son en 0.13 para la región de al menos 4 jets y 0.46 para la de al menos 6 jets.

En cuanto a la figura 28, los mejores cortes son 1 para la región de al menos 4 jets y 3 para la de al menos 6 jets.

4.7. Cortes aplicados a las distribuciones de ≥ 4 jets y ≥ 6 jets

Con los puntos de cortes obtenidos en el apartado anterior, se han representado las distribuciones de al menos 4 jets y 6 jets pero con ellos aplicados.

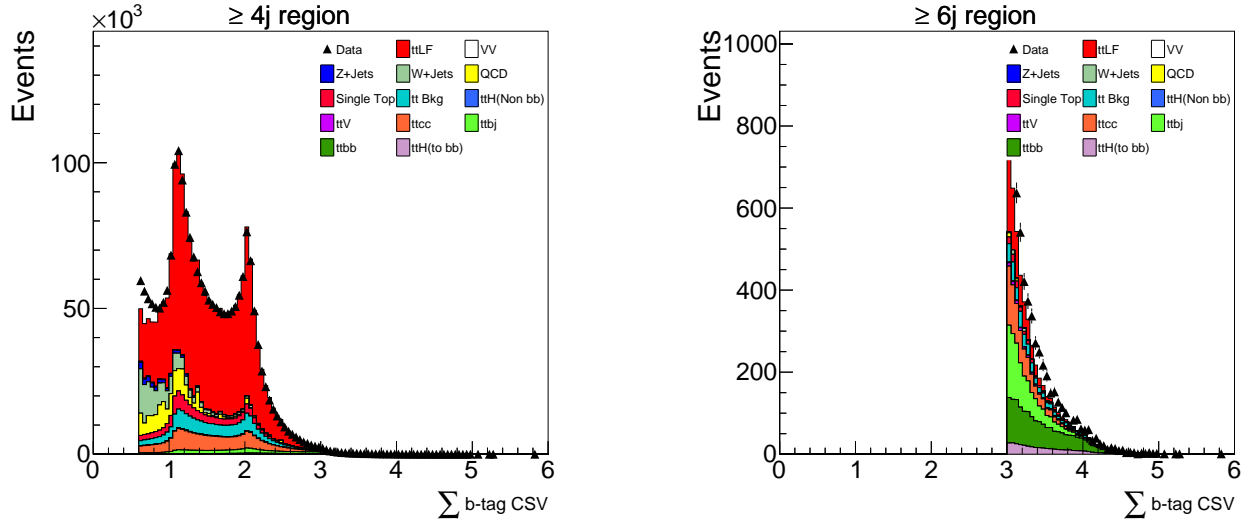


Figura 29: Representación de la variable sumatorio de b tag para las regiones indicadas.

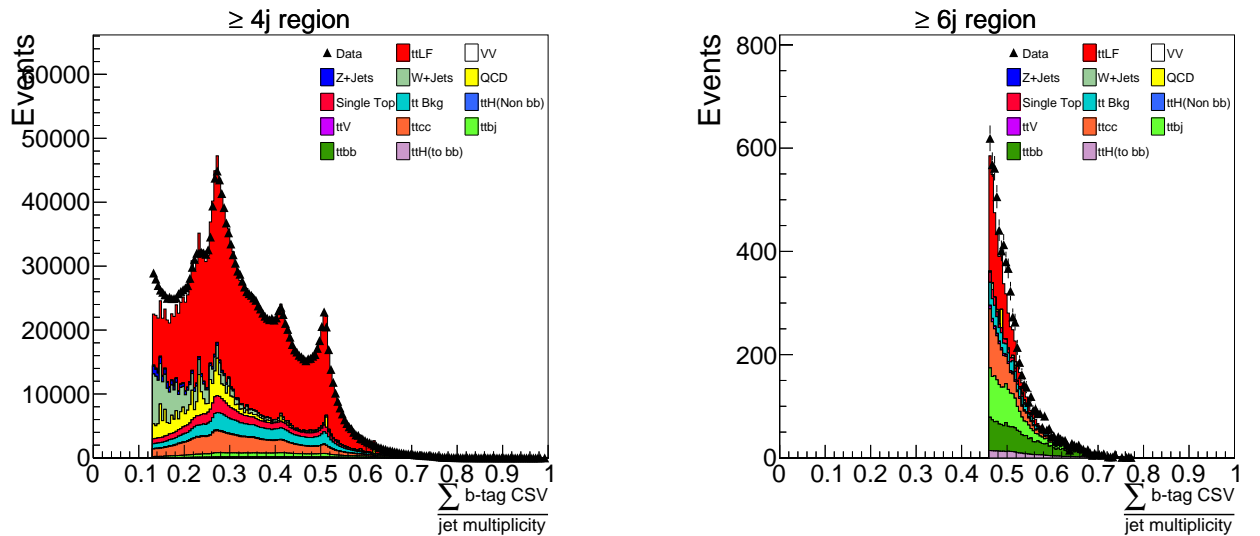


Figura 30: Representación de la variable sumatorio relativo de b-tag para las regiones indicadas.

Como en la región de ≥ 4 jets está destinada al estudio de eventos $t\bar{t}$, por lo tanto en estas gráficas hay que fijarse en la suma de los procesos $t\bar{t}b\bar{b}$ $t\bar{t}c\bar{c}$ $t\bar{t}bj$ y $t\bar{t}LF$ que serán la señal que se quiere analizar. Por otro lado, en la región ≥ 6 el interés del análisis es el estudio de los eventos $t\bar{t}b\bar{b}$.

4.8. Pureza y sección eficaz de $t\bar{t}$

A continuación se ha calculado la pureza de los eventos $t\bar{t}$ en las regiones obtenidas. Se define la pureza como

$$P_{ur} = \frac{\#EvtSgn}{\#EvtTotal} \quad (2)$$

Donde $\#EvtSgn$ es el número de eventos de la señal que se está considerando y $\#EvtTotal$ son los eventos totales. Cómo se quiere calcular la pureza del proceso $t\bar{t}$, se va a estudiar la región más favorable a este proceso, es decir, en la región de al menos 4 jets de los cuáles 2, como mínimo, son b jets y para las regiones correspondientes a las del corte aplicado para las variables definidas.

Además también se ha calculado la sección eficaz que se define como

$$\sigma(pp \rightarrow t\bar{t}) = \frac{N^{sgn}}{\xi \int \mathcal{L} dt} = \frac{N^{datos} - N^{fondo}}{\xi \int \mathcal{L} dt}$$

donde N^{sgn} es el número de eventos de la señal que se quiere estudiar y ξ es la eficiencia con la que se detecta ese proceso. Por otro lado, se encuentra $\int \mathcal{L} dt$ que es la luminosidad integrada. Para calcular este valor se necesita conocer N^{sgn} , para ello se ha expresado como la diferencia entre los datos obtenidos menos los eventos de fondo ($N^{datos} - N^{fondo}$). Solo faltaría conocer cual es el valor de la eficiencia ξ .

$$\xi = \frac{N^{medido}}{N^{total}} = \frac{N^{MC}}{\sigma_{th} \int \mathcal{L} dt}$$

De esta forma, se obtiene ξ en función de los eventos generados a partir de la simulación MonteCarlo (N^{MC}) y la sección eficaz teórica, σ_{th} , usada en la simulación. Utilizando las dos expresiones previas, se obtiene que la sección eficaz es

$$\sigma(pp \rightarrow t\bar{t}) = \sigma_{th} \frac{N^{datos} - N^{fondo}}{N^{MC}} \quad (3)$$

Donde $\frac{N^{datos} - N^{fondo}}{N^{MC}}$ se lo denomina fuerza de la señal. Con las expresiones (2) y (3) se calculan las purezas y secciones eficaces respectivamente para el proceso $t\bar{t}$, utilizando los distintos métodos de etiquetado de b-jets. Para este análisis, las secciones eficaces calculadas son estimaciones y no se han calculado las incertidumbres debido a su complejidad y a la limitación de tiempo.

	σ_{th}	$\geq 4j2b$	$\geq 4j \sum btag$	$\geq 4j \frac{\sum btag}{njet}$
$\sigma [pb]$	831.76	847	853	856
Pur	-	0.76	0.750	0.70
# Evt	-	$6.7 \cdot 10^5$	$1.41 \cdot 10^6$	$1.42 \cdot 10^6$

Tabla 4: Sección eficaz (σ), sección eficaz teórica (σ_{th}), pureza (Pur) y número de eventos ($\#Evt$) del proceso $t\bar{t}$ utilizando el método DeepCSV (región $\geq 4j2b$) y las otras dos variables construidas $\sum btag$ y $\frac{\sum btag}{njet}$

A la hora de analizar los datos, es fundamental fijarse en el balance entre la pureza y el número de eventos. La pureza obtenida para las tres regiones, $\geq 4j2b$, $\geq 4j \sum btag$ y $\geq 4j \frac{\sum btag}{njet}$ son 76 %, 75 % y 70 % respectivamente, es decir, no hay diferencias significativas. En el caso la sección eficaz, el valor teórico para este proceso es 831.76 pb, por lo que las tres estimaciones son muy buenas. Aquí se ha podido observar como para la sección eficaz y la pureza los resultados son muy similares, en cambio el número de eventos es mucho mayor (un orden de magnitud) para las nuevas variables ($\geq 4j \sum btag$ y $\geq 4j \frac{\sum btag}{njet}$). Esto conlleva una gran mejora, ya que sin sacrificar precisión en la sección eficaz y apenas reducir la pureza, el número de eventos aumenta considerablemente.

4.9. Pureza y sección eficaz de los procesos $t\bar{t}b\bar{b}$ y $t\bar{t}H$

En este apartado, se han calculado la sección eficaz y la pureza de los procesos $t\bar{t}b\bar{b}$ y $t\bar{t}H$, además del número de eventos.

Procesos	Variables	σ_{th}	$\geq 6j4b$	$\geq 6j \sum btag$	$\geq 6j \frac{\sum btag}{njet}$
ttbb	$\sigma [pb]$	4	6.2	7.1	6.9
	Pur	-	0.38	0.28	0.20
	# Evt	-	700	1400	1300
ttH	$\sigma [pb]$	0.3	0.92	1.19	1.21
	Pur	-	0.10	0.05	0.07
	# Evt	-	180	360	310

Tabla 5: Para los procesos $t\bar{t}b\bar{b}$ y $t\bar{t}H$ se muestran los valores de la sección eficaz (σ), la sección eficaz teórica (σ_{th}), la pureza (Pur) y el número de eventos ($\#Evt$) para cada región.

En el caso del proceso $t\bar{t}b\bar{b}$, la sección eficaz teórica es 4 pb por lo que el mejor resultado es el obtenido en $\geq 6j4b$, es decir, utilizando el método DeepCSV aunque son bastante similares. Puede parecer que los valores obtenidos distan mucho del valor teórico pero a causa del escaso número de eventos y una gran contribución del fondo, la incertidumbre se estima que es alrededor del 50 %. La pureza es de un 38 % mientras que para $\geq 6j \sum btag$ y $\geq 6j \frac{\sum btag}{njet}$ es del 28 % y 20 % respectivamente. Pero estas pequeñas diferencias casi no importan, porque para las nuevas variables el número de eventos es el doble lo cuál es fundamental, especialmente en estos procesos, donde estos son escasos.

Para el proceso $t\bar{t}H$ ocurre algo similar. La sección eficaz teórica es 0.3 pb por lo que el valor más cercano se obtiene mediante el DeepCSV pero los valores obtenidos para las otras

variables son similares. Hay que tener en cuenta que las incertidumbres son muy grandes, al igual que en el caso anterior. La pureza no cambia significativamente entre las regiones salvo para tal vez para $\geq 6j_{\sum_{njet}^{btag}}$ que se reduce a la mitad respecto al método DeepCSV. De todas maneras, para estas diferencias lo más importante sería el número de eventos que mediante el método DeepCSV se obtienen la mitad de ellos.

5. Conclusiones

Durante la realización de este trabajo, se ha profundizado acerca de la física del quark top y del bosón de Higgs para así poder comprender los procesos que se han analizado a lo largo del trabajo así como las propiedades más importantes del quark top. Adicionalmente, se han estudiado las diferencias entre generación y reconstrucción para así entender mejor el análisis que se ha llevado a cabo. Para el análisis de estos datos, se ha aprendido a leer la información almacenada en forma de *Tree* y el como extraer de estos la información deseada. Esto puede ser por ejemplo seleccionar el punto de trabajo (WP) o realizar la categorización de los eventos. Además, se ha aprendido C++ y *root framework* para escribir el código necesario para así poder obtener los histogramas que aparecen a lo largo del trabajo imponiendo las condiciones necesarias.

Para la realización de este trabajo, se han analizado los datos recogidos de colisiones protón-protón, en el 2017, por el CMS con una energía de centro de masas de 13 TeV.

En cuanto a los resultados obtenidos, en las tablas 4 y 5 se muestran tanto las secciones eficaces, así como la pureza y el número de eventos de los procesos $t\bar{t}$, $t\bar{t}b\bar{b}$ y $t\bar{t}H$ para las distintas variables. La región de 4j es la destinada al estudio del proceso $t\bar{t}$ inclusivo mientras que la de 6j es utilizada para los otros dos procesos.

Por un lado está el proceso $t\bar{t}$, donde claramente las nuevas variables ($\sum btag$ y $\frac{\sum btag}{njet}$) ofrecen mejores resultados que el método CSV, especialmente el sumatorio de b-tags. Esto se debe a que para las tres regiones establecidas, las secciones eficaces y purezas calculadas son muy similares, mientras que hay una enorme diferencia en el número de eventos, siendo un orden de magnitud mayor para las nuevas variables. Esto es muy importante ya que permite análisis más precisos reduciendo las incertidumbres. Las secciones eficaces son prácticamente iguales y son compatibles las tres con el valor teórico ya que como se observa en la tabla 2, las incertidumbres teóricas son de entorno a 40pb. Las purezas son también muy parecidas, apenas hay un 6 % de diferencia entre la mejor y peor medida. Por esta razón se considera que en este caso, el resultado obtenido de la variable $\sum btag$ es ligeramente mejor al de $\frac{\sum btag}{njet}$.

Para los procesos $t\bar{t}b\bar{b}$ y $t\bar{t}H$ las variables creadas son mucho más interesantes. En estos procesos, el número de eventos es muy limitado por lo que las diferencias en este aspecto son fundamentales. Además, la incertidumbre en las secciones eficaz también aumenta (en torno al 50 %) debido a los pocos eventos registrados (entre otras fuentes de incertidumbre). Desde esta perspectiva, se observa claramente como las nuevas variables son las que ofrecen mejores resultados, ya que se obtienen el doble de eventos que mediante el método DeepCSV. Adicionalmente, analizando las secciones eficaces y la pureza no se observa grandes diferencias entre las tres regiones. Las secciones eficaces son muy similares y las discrepancias con la

teórica se deben principalmente a las grandes incertidumbres mencionadas anteriormente. En cuanto a la pureza, los cambios son algo más notables que en los eventos $t\bar{t}$ inclusivos pero aún así las diferencias no son insalvables. Para el proceso $t\bar{t}H$ se puede observar como, por ejemplo, utilizando la suma de b-tags la pureza se reduce a la mitad, con lo cuál hay que tener precaución, pero si se usa la suma relativa de b-tag solo pierde un 3 % de pureza. Con el proceso $t\bar{t}b\bar{b}$ ocurriría lo mismo aunque en este caso la variable más conveniente sería la suma de b-tag.

Por todo esto, se puede concluir que las nuevas variables propuestas proporcionan ventajas respecto al método DeepCSV para la optimización de la selección de eventos, especialmente para los procesos con un reducido número de eventos. Esto se debe a que se consiguen un mayor número de eventos sin perder precisión en las estimaciones de las secciones eficaces y sin reducir considerablemente la pureza. Por lo que, al aumentar el número de eventos de los procesos que se quieran estudiar, los análisis serán más precisos.

6. Bibliografía

Referencias

- [1] Rod Nave, C. (s. f.). *Fundamental Forces*. Hyperphysics. Recuperado 4 de septiembre de 2021, de <http://hyperphysics.phy-astr.gsu.edu/hbase/Forces/funfor.html> <http://hyperphysics.phy-astr.gsu.edu/hbase/Forces/funfor.html>
- [2] Gagnon, P. (2014, 14 marzo). *The Standard Model: a beautiful but flawed theory*. Quantum Diaries. <https://www.quantumdiaries.org/2014/03/14/the-standard-model-a-beautiful-but-flawed-theory/>.
- [3] Thomson, M. (2013). *Modern particle physics*. Cambridge University Press.
- [4] ATLAS , CDF , CMS , DØ Collaboration, “First combination of Tevatron and LHC measurements of the top-quark mass”, arXiv:1403.4427.
- [5] Pivarski, J. (2013, 1 marzo). *Physics in a nutshell: Cross section*. Fermilab Today. https://www.fnal.gov/pub/today/archive/archive_2013/today13-03-01.html
- [6] Tonon, N. (2021). *CMS Top Quark Physics Summary Figures*. Twiki page. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTOPSummaryFigures>.
- [7] Czakon, M., Mitov, A. (2013). *NNLO+NNLL top-quark-pair cross sections*. Twiki page. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO#Top_quark_pair_cross_sections_at.
- [8] Najafabadi, M. M. (2006, 12 mayo). *Single Top production at LHC*. ArXiv.Org. <https://arxiv.org/abs/hep-ex/0605034>
- [9] Escobar, C. (2017, 19 septiembre). *NLO single-top channel cross sections*. Twiki page. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopRefXsec?sortcol=0;table=1;up=0#sorted_table.
- [10] Hamilton, B. (2014, 2 julio). *Semileptonic decays at LHCb* [Diapositivas]. Institut de física corpuscular. <https://indico.ific.uv.es/event/2025/contributions/816/attachments/804/931/hamilton-ichep-v3.pdf>
- [11] Sirunyan, A. M., Backhaus, M., Bäni, L., Berger, P., Chernyavskaya, N., Dissertori, G., ... CMS Collaboration. (2019). *Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV*. Journal of High Energy Physics, 2019(3), 26.
- [12] *Experiments*. (2021, 8 septiembre). CERN. <https://home.cern/science/experiments>
- [13] Mobs, E. (2019, 29 julio). *The CERN accelerator complex - 2019*. CERN Document Server. <https://cds.cern.ch/record/2684277>.
- [14] Cid Vidal, X., Cid Manzano, R. (s. f.). *Taking a closer look at LHC - CMS*. Taking a closer look at LHC. Recuperado 13 de septiembre de 2021, de https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.cms

- [15] Šalva Diblen, S. (2013). *The Compact Muon Solenoid (CMS) Experiment at CERN's Large Hadron Collider (LHC)* [Diapositivas]. CERN. https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/RetrieveFile?docid=12148&filename=CMSDetector_UNIGent.pdf
- [16] Cid Vidal, X., Cid Manzano, R. (s. f.-b). *Taking a closer look at LHC - Momentum*. Taking a closer look at LHC. Recuperado 13 de septiembre de 2021, de https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.momentum.
- [17] *Crystal Calorimeter — CMS Experiment*. (s. f.). CMS. Recuperado 13 de septiembre de 2021, de <https://cms.cern/detector/measuring-energy/crystal-calorimeter> <https://cms.cern/detector/measuring-energy/crystal-calorimeter>
- [18] Fabjan, C. W., Gianotti, F. (2003). *Calorimetry for particle physics*. Reviews of Modern Physics, 75(4), 1243. <http://cds.cern.ch/record/692252/files/ep-2003-075.pdf>
- [19] P. Giacomelli (2002). *The CMS muon detector Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 478(1), 147 - 152.
- [20] Amram, N. (2008). *Hough transform track reconstruction in the cathode strip chambers in ATLAS* (Doctoral dissertation, Tel Aviv U.). https://www.researchgate.net/publication/40618959_Hough_Transform_Track_Reconstruction_in_the_Cathode_Strip_Chambers_in_ATLAS
- [21] Asín Cruz, I. (2011). *Búsqueda de muones excitados con el detector CMS en el colisionador LHC (Search for excited muons in CMS detector in the LHC collider)*. https://eprints.ucm.es/13358/2/Ivan_Asin.pdf
- [22] Levin, H. (2013). *Quantum ChromoDynamics* [Diapositivas]. SlidePlayer. <https://slideplayer.com/slide/12399193/> <https://slideplayer.com/slide/12399193/>
- [23] Van Mulders, P. (2016, 12 abril). *Usage of b Tag Objects for 2011 Open Data Project*. Twiki page. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/BtagRecommendation2011OpenData>
- [24] Collard, C. (2017, 14 marzo). *Heavy flavor identification at CMS with deep neural networks*. Twiki page. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/BTV13TeVDPDeepCSV>